



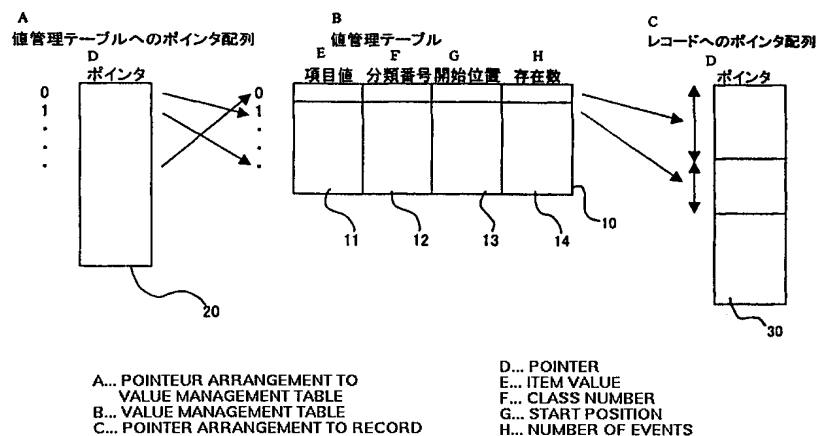
PCT

特許協力条約に基づいて公開された国際出願

(51) 国際特許分類6 G06F 17/30, 19/00		A1	(11) 国際公開番号 WO00/10103
			(43) 国際公開日 2000年2月24日(24.02.00)
(21) 国際出願番号 PCT/JP99/04300		(81) 指定国 CA, CN, JP, KR, US, 欧州特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE)	
(22) 国際出願日 1999年8月9日(09.08.99)		添付公開書類 国際調査報告書	
(30) 優先権データ 特願平10/227278 1998年8月11日(11.08.98) JP 特願平10/338133 1998年11月27日(27.11.98) JP			
(71) 出願人 ; および (72) 発明者 古庄晋二(FURUSHO, Shinji)[JP/JP] 〒221-0005 神奈川県横浜市神奈川区松見町4丁目 1101番地7 コートハウス菊名804号 Kanagawa, (JP) (74) 代理人 弁理士 窪田英一郎, 外(KUBOTA, Eiichiro et al.) 〒107-0062 東京都港区南青山5丁目12番3-501号 Tokyo, (JP)			

(54)Title: METHOD AND APPARATUS FOR RETRIEVING, ACCUMULATING, AND SORTING TABLE-FORMATTED DATA

(54)発明の名称 表形式データの検索・集計・ソート方法及び装置



(57) Abstract

A method and apparatus having a conventional function of data table, for retrieving and accumulating at a greatly improved speed a large quantity of table-formatted data represented as an arrangement of records containing item values corresponding to items of information. A value management table in which item values are stored in order of item value number corresponding to the item value belonging to a specific item and an item value number specifying information arrangement in which information for specifying item value numbers in order of record are stored in a storage. The item value number corresponding to a specific record is retrieved from the item value number specifying information arrangement, and the item value corresponding to the retrieved item value number is retrieved from the item values stored in the value management table.

従来のデータ表の機能を有すると共に大量のデータの検索・集計のスピードが大幅に向上した表形式データの検索・集計方法及び装置を提供する。本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データの検索・集計方法は、特定の項目に属する項目値に対応した項目値番号の順番に項目値が格納されている値管理テーブルと、レコードの順番に項目値番号を指定する情報が格納された項目値番号指定情報配列とを記憶装置に保持し、項目値番号指定情報配列から特定のレコードに対応した項目値番号を取得し、値管理テーブルに格納された項目値の中から、取得された項目値番号に対応した項目値を得る。

PCTに基づいて公開される国際出願のパンフレット第一頁に掲載されたPCT加盟国を同定するために使用されるコード(参考情報)

AE	アラブ首長国連邦	DM	ドミニカ	KZ	カザフスタン	RU	ロシア
AL	アルバニア	EE	エストニア	LC	セントルシア	SD	スーダン
AM	アルメニア	ES	スペイン	LI	リヒテンシュタイン	SE	スウェーデン
AT	オーストリア	FI	フィンランド	LK	スリ・ランカ	SG	シンガポール
AU	オーストラリア	FR	フランス	LR	リベリア	SI	スロヴェニア
AZ	アゼルバイジャン	GA	ガボン	LS	レソト	SK	スロヴァキア
BA	ボスニア・ヘルツェゴビナ	GB	英国	LT	リトアニア	SL	シエラ・レオネ
BB	バルバドス	GD	グレナダ	LU	ルクセンブルグ	SN	セネガル
BF	ベルギー	GE	グルジア	LV	ラトヴィア	SZ	スワジランド
BF	ブルギナ・ファソ	GH	ガーナ	MA	モロッコ	TD	チャード
BG	ブルガリア	GM	ガンビア	MC	モナコ	TG	トーゴ
BJ	ベナン	GN	ギニア	MD	モルドヴァ	TJ	タジキスタン
BR	ブラジル	GW	ギニア・ビサオ	MG	マダガスカル	TZ	タンザニア
BY	ベラルーシ	GR	ギリシャ	MK	マケドニア旧ユーゴスラヴィア	TM	トルクメニスタン
CA	カナダ	HR	クロアチア		共和国	TR	トルコ
CF	中央アフリカ	HU	ハンガリー	ML	マリ	TT	トリニダード・トバゴ
CG	コンゴ	ID	インドネシア	MN	モンゴル	UA	ウクライナ
CH	スイス	IE	アイルランド	MR	モーリタニア	UG	ウガンダ
CI	コートジボワール	IL	イスラエル	MW	マラウイ	US	米国
CM	カメルーン	IN	インド	MX	メキシコ	UZ	ウズベキスタン
CN	中国	IS	アイスランド	NE	ニジェール	VN	ヴェトナム
CR	コスタ・リカ	IT	イタリア	NL	オランダ	YU	ユーゴスラビア
CU	キューバ	JP	日本	NO	ノルウェー	ZA	南アフリカ共和国
CY	キプロス	KE	ケニア	NZ	ニュージーランド	ZW	ジンバブエ
CZ	チェッコ	KG	キルギスタン	PL	ポーランド		
DE	ドイツ	KP	北朝鮮	PT	ポルトガル		
DK	デンマーク	KR	韓国	RO	ルーマニア		

## 明 細 書

## 表形式データの検索・集計・ソート方法及び装置

5

## 発明の属する技術分野

本発明は、コンピュータのような情報処理装置を用いて大量のデータを処理するデータ処理方法及びデータ処理装置に係わり、特に、表形式のデータを検索・集計・ソートする方法及び装置に関する。

## 従来の技術

- 10 従来、大量の情報を蓄積し、蓄積された情報を検索、集計するデータ処理が行われている。これらのデータ処理は、例えば、CPUと、メモリと、周辺機器インタフェースと、ハードディスクのような補助記憶装置と、ディスプレイ及びプリンタのような表示装置と、キーボード及びマウスのような入力装置と、電源ユニットとがバスを介して接続された周知のコンピュータシステムで使用
- 15 され、特に、市場で容易に入手可能なコンピュータシステムで動作可能なソフトウェアとして提供される。上記の検索・集計等のデータ処理を行うため、特に、大量のデータを蓄積する各種データベースが知られている。大量のデータの中でも、特に、表形式で表現され得るデータを処理したいという要求は強い。図1は、処理されるべきデータを表形式で表現した例を示す図である。図1に
- 20 は、たとえば、100万人の多数の人に関して各人の性別、年齢及び職業のデータを蓄積したデータが示されている。図1において、表の横方向の行、即ち、所謂レコードは、レコード番号と、レコード番号に対応した性別、年齢及び職業の項目とからなる。表の縦方向の列は、レコード番号と、性別の項目と、年齢の項目と、職業の項目とからなる。レコード番号「0」の人は、性別が女性、
- 25 年齢が18歳、職業がプログラマであることを表す。以下の説明では、「女性」、「18歳」、「プログラマ」等のように各項目に設定されたデータを項目値と称する。また、以下の説明では、特に断らない限り、大量のデータの具

体的な例として、図 1 に示された 1 0 0 万レコードからなる表形式データを使用する。

大量のデータを効率よく検索、集計できるか否かは、大量のデータを格納する形式に依存する。従来、一般的な格納技術として図 2 (A) 及び (B) に  
5 夫々示された所謂「行単位」格納技術及び「項目単位」格納技術が知られている。

図 2 (A) および図 2 (B) は、記憶装置、たとえば、ハードディスク上のデータ格納イメージを示している。図 2 (A) の行単位格納技術の場合、各レコード番号毎に構成された性別、年齢及び職業の項目値の組がレコード番号順  
10 に、論理的アドレスが増加する順番でディスク上に格納されている。一方、図 2 (B) の項目単位格納技術の場合、各項目毎にレコード番号順に、項目値が、論理的アドレスが増加する方向にディスク上に格納されている。即ち、図 2 (B) の例では、性別に関してレコード番号「0」から「9 9 9 9 9 9」までの項目値が順番に並べられ、次に、年齢に関してレコード番号順に項目値が並  
15 べられ、それに引き続き、職業に関してレコード番号順に項目値が並べられている。

上記の従来技術の場合、全てのレコード番号の全ての項目に対する項目値が、そのまま、(レコード番号の 1 次元と、それ以外の項目値の 1 次元とからなる) 2 次元のデータ構造に格納されている。以下、このようなデータ構造を特  
20 に「データ表」と称する。従来技術の場合、蓄積されたデータを検索、集計する際には、このデータ表をアクセスすることにより行われている。

また、項目の値をそのまま項目値として格納する方法の他に、値をコード変換して、項目値としてコードを格納する方法も知られている。たとえば、性別に関して値「男性」を「0」に、値「女性」を「1」にコード変換し、「男  
25 性」又は「女性」の代わりに「0」又は「1」の値が項目値として格納される。この場合でも、コード変換されたコードが項目値としてデータ表に格納されている点に変わりはない。

上記従来技術のデータ表タイプのデータ構造を用いて格納された大量のデータを検索、集計する場合、このようなデータ表をアクセスするためのアクセス時間に起因して検索、集計の処理時間が長くなるという問題点がある。

また、データ表には少なくとも以下に述べる本質的な欠点がある。

- 5     (1) データ表はサイズが巨大化しやすく、しかも例えば項目毎等に（物理的に）分割することが困難である。例えば、性別が「男性」のレコードを抽出するとき、年齢や職業の情報は必要とされないの、性別という項目だけからなる表に分割してアクセスできれば効率的である。図2（B）に示された項目単位格納技術の場合、項目毎への分割はしやすいが、大量のデータを扱う際には
- 10    データ表のサイズが巨大化することは同じで、實際上、集計・検索のためにデータ表をメモリなどの高速な記憶装置上に展開することは困難である。

- (2) データ表は、各項目値を同時にソートした形で保持できない。例えば、図2（A）及び図2（B）に示した従来技術の場合、項目値は、性別に関して
- 15    「女性、男性、女性、．．．、女性」のようにレコード番号順に並べられている。しかし、検索・集計処理を行う場合、一般に、「女性、女性、女性、男性、．．．、男性」のように並べられている方が便宜である。ところが、データ表では、項目値は所定の配列順序、即ち、レコード番号順に並べられているので、特定の項目に関して項目値をソートすることは許容されない。そのため、従来技術の場合、検索・集計に都合の良い項目値の並べ方を選択することがで
- 20    きない。

- (3) データ表には、同一値が何度も出現する。例えば、図2（A）及び図2（B）に記載された従来のデータ表の場合、性別が『「男性」か「男」』であるレコード、即ち、レコード番号を抽出する操作の際に、「男性」という項目値が何度も出現することに起因して、「男性」という項目値と比較条件である
- 25    『「男性」か「男」』との照合を何度も行わねばならないことになる。本来的には、同一値に対して、それが該当するか否かは、1回判定すれば良いはずである。

本発明の目的は、大量のデータの検索・集計のスピードを大幅に向上させるため、従来のデータ表の機能を有すると共に、データ表に基づくデータ構造の上記問題点が解決されたデータ管理機構を設けることにより、表形式データを検索・集計・ソートする方法及びその方法を実施する装置を提供することである。

#### 発明の開示

上記目的を達成するため、本発明の表形式データを検索・集計する方法及び装置は、通常のコンピュータシステムで使用可能な新規のデータ管理機構を提案する。本発明によるデータ管理機構は原則として値管理テーブルと、値管理  
10 テーブルへのポインタ配列とを有する。

図3は、本発明の原理を説明する図であり、値管理テーブル10と値管理テーブルへのポインタ配列20とが示されている。値管理テーブル10とは、表形式データの各項目に対し、その項目に属する項目値が順序付け（整数化）された項目値番号の順番に、上記項目値番号に対応した項目値（符号11参照）  
15 と、上記項目値に関連した分類番号（符号12参照）とが格納されたテーブルである。値管理テーブルへのポインタ配列20とは、表形式データのある列（即ち、項目）の項目値番号、即ち値管理テーブル10へのポインタが表形式データのレコード番号順に格納された配列である。

値管理テーブルへのポインタ配列20と値管理テーブル10とを組み合わせ  
20 ることにより、あるレコード番号が与えられたとき、所定の項目に関する値管理テーブルへのポインタ配列20からそのレコード番号に対応して格納された項目値番号を取り出し、次に、値管理テーブル10内でその項目値番号に対応して格納された項目値を取り出すことにより、レコード番号から項目値を得ることができる。したがって、従来のデータ表と同様に、レコード番号（行）と  
25 項目（列）という座標を用いて全てのデータ（項目値）を参照することができる。

このように表形式データの項目中のある項目に対して生成された値管理テー

ブルと、値管理テーブルへのポインタ配列とを含む本発明によるデータ管理機構は、以下の説明では、特に情報ブロックと呼ぶ場合がある。

従来のデータ表は、レコードに対応した行と、項目に対応した列とからなる座標を用いて全てのデータを一体的に管理しているのに対して、本発明による

5 情報ブロックは、表形式の列、即ち、項目毎にデータを完全に分離している点に特徴がある。このように、本発明によれば、大量のデータが項目毎に分離されているので、検索・集計に必要な項目に関するデータのみをメモリ等の高速な記憶装置上に取り込むことが可能であり、その結果としてデータへのアクセス時間が短縮されるので、検索・集計の処理が高速化され、項目数が非常に多

10 いデータの場合でも、パフォーマンスを低下させることなく取り扱えるようになる。

また、本発明による情報ブロックの場合、項目値は値管理テーブルに格納され、値が存在する位置を示すレコード番号は値管理テーブルへのポインタ配列に関連付けられているので、項目値がレコード番号順に並べられている必要が

15 ない。したがって、検索・集計に適するようにデータを項目値に関してソートすることができるようになる。これにより、目的の値と一致する項目値がデータ中に存在するか否かの判定が高速に行えるようになる。その上、項目値は項目値番号と対応しているので、項目値が長いデータや文字列等であっても整数として取り扱うことができる。

20 さらに、本発明によれば、値管理テーブル 10 の全ての項目値番号は異なる項目値と対応しているので、特定の値を有する項目値を含むレコードを抽出する場合に必要とされる特定の値と項目値との比較の回数は、最大で項目値の種類の数、即ち、項目値番号の個数であり、比較演算の回数が著しく低減され、検索・集計の高速化が図られる。その際には、ある項目値が該当するかどうか

25 を調べた結果を格納する場所が必要であるが、例えば分類番号 12 をその格納場所として使用することができる。

図 4 には、項目値を格納した項目値配列 11 と、分類番号を格納した分類番

号配列 1 2 と、存在数を格納した存在数配列 1 4 とからなる値管理テーブル 1 0 を含む本発明による情報ブロックが示されている。存在数配列 1 4 には、ある項目に関する各項目値が全データ中に何個ずつ存在するかを示す数、換言すれば、所定の項目値を有するレコードの個数が格納される。このような存在数  
5 配列 1 4 を値管理テーブル 1 0 に準備しておくことにより、検索・集計の際に必要とされる「どのようなデータが（いくつ）存在するか」という情報を直ちに得ることができるようになり、検索・集計の高速化が図れる。

図 5 には、値管理テーブル 1 0 と、値管理テーブルへのポインタ配列 2 0 と、レコードへのポインタ配列 3 0 とからなる情報ブロックが示されている。レコードへのポインタ配列 3 0 とは、項目値番号、即ち、項目値毎に、その項目値  
10 を有するレコードへのポインタ（レコード番号に対応する）が格納された配列である。各項目値毎にレコードへのポインタ配列 3 0 に格納されるポインタの個数は、値管理テーブル 1 0 の存在数配列 1 4 に格納された数と一致する。また、レコードへのポインタ配列 3 0 内で、各項目値毎のポインタの群の先頭の  
15 アドレスを指定する開始位置を格納した開始位置配列 1 3 を値管理テーブル 1 0 に設けてもよい。このようなレコードへのポインタ配列 3 0 を情報ブロックに設けることにより、所定の項目に関して、特定の項目値を有するレコードの集合を高速に抽出できるようになる。値管理テーブル 1 0 にレコードへのポ  
インタ配列 3 0 に格納されるポインタの存在数（符号 1 4 参照）及びその開始位  
20 置（符号 1 3 参照）が設定されているので、集計時に利用可能な値、個数が情報ブロックにそのまま存在している点が有利である。

次に、本発明による表形式データの検索・集計方法について説明する。なお、以下の説明では、個別項目情報は上記の「情報ブロック」を表し、項目値番号指定情報配列は上記の「値管理テーブルへのポインタ配列」を表し、レコード  
25 識別情報配列は上記の「レコードへのポインタ配列」を表す。

本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データから、特定の項目及び特定のレコードに対応した項目値を得



る方法は、

特定の項目に属する項目値に対応した項目値番号の順番に、項目値が格納されている値管理テーブルと、レコードの順番に項目値番号を指定する情報が格納された項目値番号指定情報配列とを記憶装置に保持し、

- 5      項目値番号指定情報配列から特定のレコードに対応した項目値番号を取得し、  
値管理テーブルに格納された項目値の中から、取得された項目値番号に対応した項目値を得るステップを含む。

- また、本発明による項目値を得る方法によれば、項目値番号に対応した項目値を分類するための分類番号が、項目値番号に対応して値管理テーブルに格納  
10      され、項目値番号に対応した項目値を得る際に分類番号が参照される。

本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データから検索条件に適合する項目値を含むレコードを検索する単項目検索方法は、

- 検索条件と関連した項目に属する項目値に対応した項目値番号の順番に、項  
15      目値が格納されている値管理テーブルと、レコードの順番に項目値番号を指定する情報が格納された項目値番号指定情報配列と、同じ項目値番号に対応した上記レコードの識別情報が項目値番号毎に別個に格納されたレコード識別情報配列とを含み、値管理テーブルはレコード識別情報配列の中で同じ項目値番号に対応したレコードの識別情報が格納されている場所を示すレコード識別情報  
20      指定情報が項目値番号毎に格納されている、個別項目情報を記憶装置に保持し、

値管理テーブルに格納された項目値の中で、検索条件に適合する項目値と関連した項目値番号に対応して格納されたレコード識別情報指定情報を用いて、レコード識別情報配列から検索条件に適合するレコードの識別情報を取り出すステップを含む。

- 25      また、本発明による複数項目検索方法は、

本発明の単項目検索方法に従って得られた上記検索条件に適合するレコードの識別情報に基づいて上記検索条件に適合するレコードを指定する結果集合を

記憶装置に保持し、

別の検索条件と関連した項目に関する別の個別項目情報を選択し、

別の個別項目情報の項目値番号指定情報配列から、上記結果集合で指定された上記検索条件に適合したレコードの識別情報に対応した項目値番号を取り出し、

別の個別項目情報について取り出された項目値番号と関連した項目値が、別の検索条件と適合するか否かを判定し、

別の個別項目情報について項目値番号と関連した項目値が、別の検索条件と適合する場合に、項目値番号に対応した上記レコードの識別情報を、別の検索条件に適合するレコードの識別情報として取り出すステップを含む。

或いは、複数項目検索方法を変形することにより、いわゆるOR検索を実現することもできる。より詳細には、検索条件に適合するレコードの識別情報に基づいて検索条件に適合するレコードを指定する結果集合を記憶装置に保持し、

他の検索条件と関連した項目に関する他の個別項目情報を記憶装置に保持し、  
他の個別項目情報の値管理テーブルに格納された項目値の中で、検索条件に適合する項目値と関連した項目値に対応するレコード識別情報指定情報を用いて、レコード識別情報配列から、他の検索条件に適合するレコード識別情報を取り出し、検索条件に適合するレコードを指定する他の結果集合を記憶し、

必要な場合に、さらに他の検索条件に関して、さらに他の個別項目情報について、個別項目情報の記憶、レコード識別情報の取り出し、および、さらに他の結果集合の記憶を繰り返して、得られた結果集合から重複したレコードを除去した最終的な結果集合を得る。

本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データを項目値毎に集計する方法は、

$n$ が1以上の整数を表す場合に、集計に利用される $n$ 個の項目について、当該項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、レコードの順番に項目値番号を指定する情報が格納さ

れた項目値番号指定情報配列とにより表された  $n$  個の個別項目情報を記憶装置に保持し、

$i$  が  $1 \leq i \leq n$  なる整数を表すとき、 $i$  番目の個別情報項目について、項目値番号の総数が  $N_i$  で表され、 $k_i$  が  $0 \leq k_i \leq N_i - 1$  なる整数を表し、 $M$  が 1  
5 以上の整数を表し、 $m$  が  $1 \leq m \leq M$  なる整数であるとき、 $N_1 \times N_2 \times \dots \times N_i \times \dots \times N_n$  のサイズを持つ、 $n$  次元の、 $M$  個のデータ空間の各要素  $P_m$  ( $k_1, k_2, \dots, k_i, \dots, k_n$ ) を初期化し、

$n$  個の個別情報項目について、 $j$  が  $0 \leq j \leq (\text{レコードの総数} - 1)$  なる整数を表すとき、各項目値番号指定情報配列の  $j$  番目に格納された夫々の項目値  
10 番号を取り出し、 $i$  番目の個別情報項目から取り出された項目値番号が  $q_i$  で表されるとき、データ空間の要素  $P_m$  ( $q_1, q_2, \dots, q_i, \dots, q_n$ ) を特定し、

特定された要素  $P_m$  ( $q_1, q_2, \dots, q_i, \dots, q_n$ ) の値を処理するステップを含む。

15 また、本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データを項目値の分類毎に集計する方法は、

$n$  が 1 以上の整数を表し、集計に利用される  $n$  個の項目について、項目に属する項目値に対応した項目値番号の順番に、項目値及び項目値の分類番号が格納されている値管理テーブルと、レコードの順番に、項目値番号を指定する情  
20 報が格納された項目値番号指定情報配列とにより表された  $n$  個の個別項目情報を記憶装置に保持し、

$i$  が  $1 \leq i \leq n$  なる整数を表すとき、 $i$  番目の個別情報項目について、項目値番号又は分類番号の何れか一方の番号の総数が  $N_i$  で表され、 $k_i$  が  $0 \leq k_i \leq N_i - 1$  なる整数を表し、 $M$  が 1 以上の整数を表し、 $m$  が  $1 \leq m \leq M$  なる整  
25 数であるとき、 $N_1 \times N_2 \times \dots \times N_i \times \dots \times N_n$  のサイズを持つ、 $n$  次元の、 $M$  個のデータ空間の各要素  $P_m$  ( $k_1, k_2, \dots, k_i, \dots, k_n$ ) を初期化し、

n個の個別情報項目について、 $j$ が $0 \leq j \leq (\text{レコード総数} - 1)$ なる整数を表すとき、各項目値番号指定情報配列の $j$ 番目に格納された夫々の項目値番号を取り出し、 $i$ 番目の個別情報項目から取り出された項目値番号又は上記 $i$ 番目の個別情報項目の値管理テーブルに該項目値番号と対応して格納された分類番号が $q_i$ で表されるとき、上記データ空間の要素 $P_m (q_1, q_2, \dots, q_i, \dots, q_n)$ を特定し、

特定された要素 $P_m (q_1, q_2, \dots, q_i, \dots, q_n)$ の値を処理するステップを含む。

特に、本発明による個数を集計する方法によれば、 $M = 1$ であり、特定された要素 $P_m$ の値を処理するステップでは、要素 $P_m$ の現在の値に1を加える。

また、本発明による統計的な量を集計する方法によれば、特定された要素 $P_m$ の値を処理するステップは、 $M$ 個の要素 $P_m$ の中の少なくとも1個の要素 $P_m$ に対して、

記憶装置に保持された別の個別項目情報について項目値番号指定情報配列の $j$ 番目に格納された項目値番号を取得し、

別の個別項目情報の値管理テーブルに格納された項目値の中から、取得された項目値番号に対応した項目値を取得し、

要素 $P_m$ の現在の値と上記取得された項目値を組み合わせて、要素 $P_m$ の値を更新する。

本発明によれば、項目値番号を指定する情報は項目値番号自体でもよい。

或いは、あるレコードの一つの項目に対し複数の項目値が割り当てられないいわゆるマルチアンサ項目を実現するため、本発明によれば、上記項目値番号を指定する情報は、各項目値番号毎に1ビット以上が割り当てられ、指定の有無が設定される2進値でも構わない。

また、本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データを検索・集計する装置は、

項目に属する項目値に対応した項目値番号の順番に、項目値が格納されてい

る値管理テーブルと、レコードの順番に、項目値番号を指定する情報が格納された項目値番号指定情報配列とを保持する記憶装置と、

記憶装置に保持された項目値番号指定情報配列から、特定のレコードに対応した項目値番号を取得する手段と、

- 5 記憶装置に保持された値管理テーブルに格納された項目値の中から、取得された項目値番号に対応した項目値を得る手段とを有することを特徴とする。

また、本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データを検索・集計するプログラムを記録した記憶媒体は、

- 10 項目に属する項目値に対応した項目値番号の順番に、項目値が格納されている値管理テーブルと、レコードの順番に、項目値番号を指定する情報が格納された項目値番号指定情報配列とを記憶装置に保持させるステップと、

記憶装置に保持された項目値番号指定情報配列から、特定のレコードに対応した項目値番号を取得させるステップと、

- 15 記憶装置に保持された上記値管理テーブルに格納された項目値の中から、上記取得された項目値番号に対応した項目値を得させるステップとを有することを特徴とするプログラムを記録した記憶媒体である。

- 本発明は、更に、情報の項目に対応した項目値を含むレコードを指定するレコード識別情報、例えば、レコード番号の配列を、項目に関して並べ換えるレコードのソート方法を提案する。本発明のソート方法によれば、レコード毎に、
- 20 レコード識別情報とある項目の項目値に対応した項目値番号とが関連付けられた値管理テーブルへのポインタ配列が形成される。次に、項目値番号毎に、レコード識別情報の並び換え後の格納場所が値管理テーブル中に定義される。配列から上記レコード識別情報が順番に取り出され、取り出されたレコード識別情報に対応した上記項目値番号が判定され、取り出されたレコード識別情報は
- 25 判定された項目値番号に対応して定義された格納場所に格納され、レコード識別情報が格納される格納場所が、次のレコード識別情報が格納され得るように更新される。

本発明にかかるソート方法の好ましい実施態様においては、検索条件と関連した項目に属する項目値に対応した項目値番号の順番に、項目値が格納されている値管理テーブルと、レコードの順番に項目値番号を指定する情報が格納された項目値番号指定情報配列とを含む個別項目情報を記憶装置に保持し、値管理5 テーブルが、さらに、項目値番号ごとに、レコード識別情報配列において相等しい項目値番号に関連する一以上のレコード識別情報を格納した領域を示すレコード識別情報指定情報を含み、格納するステップが、レコード識別情報指定情報にしたがって、レコード識別情報を、格納場所に格納するように構成されている。

- 10 さらに、本発明の目的は、上記方法を実現するための装置、方法にかかるプログラムを格納したコンピュータにより読み出し可能な記憶媒体、或いは、当該方法にかかる、コンピュータによりロード可能なプログラムプロダクトにより達成される。

#### 図面の簡単な説明

- 15 本発明の目的および他の目的は、添付図面とともに実施例を参照することによりより明らかになるであろう。ここに、

図1は、一般的な表形式データの説明図である。

図2（A）及び図2（B）は、従来技術による表形式データの格納技術の説明図である。

- 20 図3は、本発明の原理説明図である。

図4は、本発明による情報ブロックの説明図である。

図5は、本発明による情報ブロックの説明図である。

図6は、本発明の実施例で使用される「性別」に関する情報ブロックの説明図である。

- 25 図7は、本発明の実施例で使用される「年齢」に関する情報ブロックの説明図である。

図8は、本発明の実施例で使用される「性別」に関する情報ブロックの説明

図である。

図9は、本発明の第1実施例による単一項目内の検索方法の動作フローチャートである。

図10は、本発明の第1実施例による情報ブロックの説明図である。

5 図11は、本発明の第1実施例による情報ブロックの説明図である。

図12は、本発明の第2実施例による複数項目のAND検索方法の動作フローチャートである。

図13は、本発明の第2実施例による情報ブロックの説明図である。

図14は、本発明の第2実施例による情報ブロックの説明図である。

10 図15は、本発明の第3実施例によるビットフラグを用いた複数項目論理演算検索方法の説明図である。

図16は、本発明の第3実施例によるビットフラグを用いた複数項目論理演算検索方法の説明図である。

図17は、本発明の第5実施例による集計方法の動作フローチャートである。

15 図18は、本発明の第6実施例の概念的な説明図である。

図19は、本発明の第6実施例の動作フローチャートである。

図20は、本発明の第7実施例によるクロス集計の動作フローチャートである。

図21は、本発明の第8実施例による情報ブロックの説明図である。

20 図22は、本発明の第9実施例によるクロス集計の動作フローチャートである。

図23(A)及び図23(B)は、クロス集計表の概念説明図である。

図24は、 マルチアンサ形式の項目の説明図である。

25 図25は、本発明の第10実施例によるマルチアンサ項目対応型の情報ブロックの説明図である。

図26は、本発明の第11実施例による特殊値の取扱方法の説明図である。

図27は、本発明の第12実施例による複数項目検索方法の動作フローチャ

ートである。

図28は、本発明の一実施例による表形式データの検索・集計システムの構成図である。

図29は、情報ブロックの構築方法の説明図である。

5 図30は、データ投入の準備と初期化の説明図である。

図31は、データ投入の第1のパスの説明図である。

図32は、データ投入の第2のパスの説明図である。

図33は、データ投入の第3のパスの説明図である。

図34は、データ投入の第3のパスの説明図である。

10 図35は、データ投入の第3のパスの説明図である。

図36は、情報ブロックに対するデータ追加の説明図である。

図37は、本発明の他の実施例による情報ブロックの構造を示す図である。

図38は、本発明の第13実施例によるソートの初期状態の説明図である。

15 図39は、本発明の第13実施例によるソートの第1ステップの説明図である。

図40は、本発明の第13実施例によるソートの第2ステップの説明図である。

図41は、本発明の第13実施例によるソートの最終状態の説明図である。

図42は、部分集合に対するソートの説明図である。

20 図43は、部分集合に対するソートの後処理の説明図である。

図44は、検索・集計テストに使用した100万件のデータの説明図である。

図45は、100万件のデータの検索・集計テストの測定結果を示す図である。

25 図46(A)および図46(B)は、本発明の第2実施例の変形にかかる、複数項目のOR検索処理を示すフローチャートである。

図47は、本発明の第3実施例にかかる検索処理を示すフローチャートである。



図48は、本発明の第4実施例にかかる集計処理を示すフローチャートである。

図49は、本発明の第7実施例にかかる集計処理を示すフローチャートである。

5 図50は、本発明の第13実施例にかかるソート処理を示すフローチャートである。

#### 発明の好ましい実施形態の説明

以下、本発明がより良く理解できるように図1に示した表形式データをデータの例として使用して、本発明にかかる検索方法、集計方法およびソート方法  
10 の種々の実施例を詳細に説明する。図1の例に示されるデータは、「性別」、「年齢」及び「職業」の項目を含むので、情報ブロックとして、図6ないし図8に夫々示されているように、「性別」に関する情報ブロック、「年齢」に関する情報ブロック及び「職業」に関する情報ブロックが得られる。以下の実施例の説明では、これらの情報ブロックが得られた状況を想定している。なお、  
15 情報ブロックを構築する手法の一例は後述するが、本発明は情報ブロックを構築する方法によって制限されないことに注意する必要がある。

後述するように、本発明の実施例にかかる表形式データの検索・集計装置は、図28に示す構成を備えている。図28に示すように、表形式データの検索・集計装置は、通常のパーソナルコンピュータのようなコンピュータシステムにより実現される。このコンピュータシステムは、プログラムを実行することにより、システム全体および個々の構成部分を制御するCPU100、プログラム等を記憶するROM(Read Only Memory)110、ワークデータなどを記憶するRAM(Random Access Memory)120、ハードディスク記憶装置130、表示装置140およびキーボードやマウスからなる入力装置150を備えている。これらCPU100、ROM110、RAM120等は、バス160を介して相互に接続  
20 されている。また、バスには、CD-ROMをアクセスするためのCD-ROMドライバ（図示せず）、外部ネットワーク（図示せず）と接続された外部端子と

の間に設けられたインタフェース（図示せず）などが接続されていてもよい。

本実施の形態にかかる、表形式データを検索・集計（場合によってはこれに加えてソート）するプログラムは、CD-ROM（図示せず）に収容され、CD-ROMドライバ（図示せず）に読み取られても良いし、ROM110に予め記憶  
5 されていても良い。また、いったんCD-ROMから読み出したものを、ハードディスク記憶装置130の所定の領域に記憶しておいても良い。或いは、上記プログラムは、ネットワーク、外部端子およびI/F（何れも図示せず）を経て外部から供給されるものであっても良い。

また、上記検索・集計装置において、表形式データの検索・集計（場合によつてはこれに加えてソート）処理を実行するために、後述するように、表形式データに基づき、所定のデータ形式の情報ブロックを生成する必要がある。この情報  
10 ブロック生成プログラムも同様に、CD-ROMに収容され、ROM110に記憶され、或いは、ハードディスク記憶装置130に記憶されても良い。或いは、これらプログラムは、ネットワーク（図示せず）を介して、外部から供給されても良いことはいうまでもない。また、本実施の形態において、上記情報ブロックを生成する情報ブロック生成プログラムにて生成されたデータ（情報ブロック）は、RAM120、或いは、ハードディスク記憶装置130の所定の領域に記憶  
15 される。以下に、「年齢」の項目値が「16歳」又は「19歳」であるレコードを検索する場合に関して、本発明の第1実施例による単一項目内の検索方法を説明する。図9は、単一項目内の検索方法の動作フローチャートである。これは、CPU100が、上述した手順にて取得し、所定の位置に記憶した検索プログラムを実行することにより実現される。

最初に、表形式データに関する情報ブロックの中から、図7に示された「年齢」に関する情報ブロックを特定情報ブロックとして選択する（ステップ10  
25 0）。

次に、特定情報ブロックの値管理テーブルの中で項目値が上記の検索条件である「16」又は「19」に合致する行の分類番号を「1」にセットし、それ

以外の行の分類番号を「0」にセットする（ステップ102）。図10に示されるように本例の場合、項目値番号「0」及び項目値番号「3」に対応する行の分類番号が「1」にセットされる。

- 続いて、分類番号が「1」にセットされている行（即ち、項目値番号「0」及び「3」が付与されている行）に対応した開始位置及び存在数を、ポイント  
5 取り出し情報として取得する（ステップ104）。本例の場合、項目値番号「0」に対応する開始位置は「0」であり、その存在数は「45898」である。その一方、項目値番号「3」に対応する開始位置は「238137」であり、その存在数は「189653」である。
- 10 レコードへのポイント配列の中で、上記の開始位置から存在数で指定された個数分のポイントを取り出すことにより、上記検索条件に合致したレコードへのポイントを表すレコード番号が取り出される（ステップ106）。本例の場合、図10に示されているように、項目値番号「0」に対応したレコードへの  
ポイントは、レコードへのポイント配列の開始位置「0」、即ち、先頭から、  
15 45898個目までの場所に格納され、項目値番号「3」に対応したレコードへのポイントは、レコードへのポイント配列の2383137番目から189653個分の場所に格納されていることが分かる。例えば、図1の表形式データを参照するに、最後のレコード番号「999999」のレコードに対応する  
「年齢」は「16歳」であるので、図11に示される如く、レコードポイント  
20 への配列内で項目値番号「0」、即ち、「16歳」の「年齢」に対応して格納されたポイントの中で最後のポイントは、「999999」である。

最後に、後の処理で利用するため、取り出されたレコード番号の配列を結果集合として作成、保持する（ステップ108）。

- 本発明においては、上述したような単一項目による検索のみでなく、複数項  
25 目のAND検索を実現することもできる。以下、本発明の第2実施例にかかる複数項目のAND検索方法を説明する。本例では、「年齢」が「16歳」又は「19歳」であるという第1の検索条件と、「職業」が「学生」であるという

第2の検索条件を満たすレコードの集合を得る場合を考える。図12は、複数項目のAND検索方法の動作フローチャートである。

既に説明したように、第1の項目である「年齢」に関する情報ブロックである第1の特定情報ブロックについて、「年齢」が「16歳」又は「19歳」であるレコードの結果集合は第1実施例にかかる処理に従って得られる（ステップ120）。したがって、このステップ120の処理は、図9に示すものと略対応する。

次に、図8に示されている第2の項目である「職業」に関する情報ブロックである第2の特定情報ブロックとして選択する（ステップ122）。

10 次に、第2の特定情報ブロックの値管理テーブルの中で、項目値が上記の検索条件である「学生」に合致する行の分類番号を「1」にセットし、それ以外の行の分類番号を「0」にセットする（ステップ124）。図13に示されるように本例の場合、項目値番号「0」に対応する行の分類番号が「1」にセットされ、それ以外の行は「0」にセットされる。

15 次に、第1の検索条件に関する結果集合から、レコードへのポインタを表すレコード番号を順次取り出す（ステップ126）。たとえば、図14に示されるように、レコード番号「999999」が取り出される。

次に、第2の特定情報ブロックに関して、値管理テーブルへのポインタ配列から上記第1の検索条件について得られたレコード番号に対応した項目値番号を取り出す（ステップ128）。たとえば、図14に示されるように、レコード番号「999999」に対応した項目値番号「0」が取り出される。

次に、第2の特定情報ブロックに関して取り出された項目値番号に対応する分類番号が「1」にセットされているか否かを判定する（ステップ130）。たとえば、図14に示されているように、項目値番号「0」に対応する分類番号は「1」にセットされていることがわかる。

分類番号が「1」にセットされている場合、値管理テーブルへのポインタ配列内で、分類番号が「1」にセットされている当該項目値番号を指すポインタ

が格納されている場所に対応したレコードへのポインタ、例えば、レコード番号を最終的な結果集合に追加する（ステップ132）。たとえば、図14に示されているように、レコード番号「999999」が最終的な結果集合に追加される。

- 5 分類番号が「0」である場合には、最終的な結果集合は更新されない。

なお、当業者であれば容易に分かるように、上記本発明の第2実施例による複数項目AND検索方法は、AND以外の検索、例えば、複数項目OR検索方法に変形することが可能である。図46（A）は、複数項目のOR検索方法の処理の一例を示すフローチャートである。この処理もCPU100が所定領域  
10 に格納されたプログラムを実行することにより実現される。図46に示すように、まず、第1の検索条件について結果集合を取得した（ステップ4601）後に、第2の検索条件にかかる情報ブロックを選択する（ステップ4602）。次いで、この情報ブロックに関して第2の検索条件について分類番号をセットする（ステップ4603）。そして、第1の検索条件の結果集合に格納された  
15 レコード番号をスキップしながら、第2の特定情報ブロックについての値管理テーブルへのポインタ配列を順次走査する（ステップ4604）。より詳細には、第2の検索条件に関して分類番号が「1」となっているレコード番号を参照し、このレコード番号が、第1の検索条件による結果集合中に見出されるか否かが判断される（ステップ4611～4615参照）。第1の検索条件によ  
20 る結果集合中にその番号が見出されなかった場合には、結果集合にその番号が追加される（ステップ4614）。このような処理が終了した後に、第1の検索条件の結果集合に格納されたレコード番号と、第2の情報ブロックについて分類番号がセットされた項目値番号に属するレコード番号とを併せて第2の結果集合として生成し（ステップ4615）、これを出力することができる。

- 25 或いは、図46（B）に示す処理を実行してもよい。この例では、第1の特定情報に関する検索条件により第1の結果集合が得られた後に、これとは独立して、第2の特定情報に関する第2の検索条件により第2の結果集合を得て

(ステップ4621～ステップ4624)、第1の結果集合と、第1の結果集合(ステップ4625参照)との間で、ビットマップを用いたOR処理を施して(ステップ4626)、これにより新たな結果集合が作成される(ステップ4627)。なお、図46(B)の処理において、ステップ4602、4603は、図46(A)のステップ4621、4622に対応し、ステップ4625は、図46(A)のステップ4601に対応する。

図15及び図16は本発明の第3実施例によるビットフラグを用いた複数項目論理演算検索方法の説明図であり、上記本発明の第2実施例による検索と同一の検索条件で検索を行う場合を示している。ビットフラグを用いた複数項目論理演算検索とは、検索条件が項目毎の検索条件の間の論理演算によって表されるような検索を意味する。この場合、単一項目の検索により得られる結果集合を、第1実施例の如く、レコード番号の配列により構成するのではなく、結果集合をビットフラグの形式で構成する方が有利である。即ち、図47に示す処理にしたがって、全レコードに対し1ビットずつを割り当て、各レコードが検索条件に適合しているか否かをビットの値「1」又は「0」によって表わすような結果集合を生成する。より詳細には、まず、他の実施例と同様に、検索条件にかかる項目値を含む情報ブロックが選択され(ステップ4701)、検索条件に合致する行の分類番号を「1」にセットする(ステップ4702)。次いで、各レコードについて、対応する分類番号を参照して、結果集合に格納すべきビット値を決定する(ステップ4703～ステップ4707)。結果集合をこのように構成することにより、各項目毎の結果集合のサイズが表形式データのレコード数と対応するので、各項目毎の結果集合のサイズが一致し、その結果として、結果集合中の要素間の論理演算、例えば、AND、OR、XORを簡単に行うことができる。

本例では、図15に示された結果集合Aと図16に示された結果集合BをAND条件で結合することにより所望の検索結果集合がビットフラグ形式で得られる。また、得られたビットフラグ形式の検索結果集合は、レコードへのポイ

ンタ配列形式の結果集合に変換することにより、上記本発明の第2実施例による複数項目検索方法と組み合わせて用いることができる。

次に、本発明にかかる種々の表形式データの集計方法につき説明を加える。本発明の第4実施例による集計方法は、特定の項目に関して特定の項目値を有するレコード数をカウントする。本発明の第4実施例では、特に、図1の表形式データに関して、項目「性別」について「男性」という項目値及び「女性」という項目値を有するレコードの数を夫々カウントする場合を考える。既に、図6に示しているように、本発明の好ましい実施例によれば、「性別」に関する情報ブロックには、項目値「男性」を含むレコードの存在数（その値は「632564」である）、および、項目値「女性」を含むレコードの存在数（その値は「367436」である）が格納されているので、単純なレコード数の集計は情報ブロック中の存在数配列を参照することによって直ちに得ることができる。

また、単一項目の検索方法に関連して説明した分類番号を、上記本発明の第4実施例による集計方法と組み合わせることにより、更に複雑な条件の場合にもレコード数のカウントを簡単に行うことができる。例えば、本発明の第1実施例で説明した「年齢」の項目値が「16歳」又は「19歳」であるレコードを検索する単一項目検索方法において、分類番号が「1」にセットされた項目値番号に対応した存在数を合計することにより、検索条件に適合したレコード数を集計することが可能である。このように、分類番号を使用することにより、値管理テーブルのサイズが大きい場合、或いは、複雑な条件が与えられた場合であっても、存在数を効率的に求めることができるようになる。より一般には、図48に示すように、分類番号が「1」にセットされた項目値番号を見出し（ステップ4802参照）、対応する存在数を加算していけばよい（ステップ4803参照）。

次に、本発明の第5実施例につき説明を加える。この実施例では、「男性」の平均年齢が算出されている。平均年齢は、（男性の「年齢」の総和）／（男

性の存在数)により求めることができ、また、男性の存在数は、上記の本発明の第4実施例で説明した集計方法により得ることができる。従って、本実施例は、男性の「年齢」の総和を求める問題に帰着される。図17は本発明の第5実施例の動作を示すフローチャートである。この処理も、他の実施例と同様に、

5 CPU100が所定の領域に記憶されたプログラムを読み出し、これを実行することにより実現される。

最初に、図6に示された「性別」に関する情報ブロックを第1の情報ブロックとして選択し(ステップ140)、第1の特定情報ブロックの値管理テーブルの中で、項目値「男性」に対応した項目値番号「0」を検出する(ステップ

10 142)。次に、項目値番号「0」に対応した存在数が「632564」であることから男性の総数が632564人であることを判定し(ステップ144)、また、項目値番号「0」に対応した開始位置が「0」であることから、性別が男性であるレコードへのポインタは、レコードへのポインタ配列の先頭から632564個目までの場所に格納されていることを判定し、このレコードへのポインタの一覧、即ち、レコード番号の配列を結果集合として保持する

15 (ステップ146)。

次に、図7に示された「年齢」に関する情報ブロックを第2の特定情報ブロックとして選択し(ステップ148)、第2の特定情報ブロックの値管理テーブルへのポインタ配列から、第1の特定情報ブロックに関する結果集合で指定

20 されたレコードに対応した項目値番号を取り出し(ステップ150)、取り出された項目値番号と関連した項目値、即ち、「年齢」を取り出す(ステップ152)。最後に、取り出された「年齢」を順次加算して年齢の総和を求め(ステップ154)、上記結果集合で指定されたレコードを全て処理するまでステップ150、152、154を繰り返す(ステップ156)。得られた年齢の

25 総和を存在数で除算することにより年齢の平均が求められる(ステップ158)。

次に、本発明の第6実施例につき説明を加える。この例では、学生の男性の



平均年齢、及び、学生の女性の平均年齢を求めている。。図18は本発明の第6実施例を概念的に説明する図であり、図19は本発明の第6実施例にかかる動作を示すフローチャートである。

本実施例の集計は、最初に、第1の情報ブロックとして「職業」に関する情報ブロックに関する情報ブロックを選択し（ステップ170）、「職業が学生である」という検索条件を用いて、全てのレコードの中から「職業が学生である」レコードを含む結果集合を作成する（ステップ172）。

次に、「性別」に関する情報ブロックを第2の情報ブロックとして選択するとともに、「年齢」に関する情報ブロックを第3の情報ブロックとして選択し（ステップ174）、結果集合の先頭から順次レコードへのポインタを取り出す（ステップ176）。

取り出されたレコードへのポインタを用いて、第2の情報ブロックの値管理テーブルへのポインタ配列を参照して取り出されたレコードへのポインタに対応した性別を獲得し、また、第3の情報ブロックの値管理テーブルへのポインタ配列を参照して取り出されたレコードへのポインタに対応した年齢を取り出す（ステップ178）。取り出された性別の男女別の総数を1ずつ増加し、取り出された年齢の合計を男女別に計算する（ステップ180）。

結果集合の全てのレコードへのポインタが処理されたか否かを検査し（ステップ182）、全てのレコードへのポインタが処理された場合に、男女別に年齢の合計を総数で除算することにより学生の男女別の平均年齢を計算する（ステップ184）。

次に、図20に示された動作フローチャートを参照して、本発明の第7実施例につき説明を加える。この実施例では、いわゆるクロス集計を実現している。なお、図20に示す処理プログラムも、他の実施例と同様に、CPU100により読み出され、実行される。本実施例では、図1の表形式データに関してレコード全体を対象として、性別／職業別の存在数を求める場合を考える。

集計に利用される「性別」及び「職業」の2個の項目について、夫々の値管

理テーブルと、項目値番号指定情報配列、即ち、値管理テーブルへのポインタ配列とにより表された2個の個別項目情報、即ち、第1及び第2の情報ブロックを記憶装置に保持する（ステップ190）。記憶装置は、例えば、メモリ、仮想記憶、メモリマップドファイル等の形で実現される。

- 5 性別に関する第1の情報ブロックについて、図6に示される如く性別の項目値番号の総数は「2」であり、また、職業に関する第2の情報ブロックについて、図8に示される如く、職業の項目値番号の総数は「4」であることが分かる。そこで、集計データを格納する空間として、初期的な $2 \times 4$ （2行4列）の2次元配列を生成する（ステップ192）。
- 10 第1及び第2の情報ブロックについて、夫々の値管理テーブルへのポインタ配列の先頭から順に夫々の項目値番号 $q_1$ 及び $q_2$ を取り出し、これにより、2次元配列の1個の要素 $P(q_1, q_2)$ を特定し（ステップ194）、特定された要素 $P(q_1, q_2)$ の値を1だけ増加させる（ステップ196）。

- 値管理テーブルへのポインタ配列から全ての項目値番号（即ち、レコード総  
15 数個の項目値番号）が取り出されたか否かを検査し（ステップ198）、未だ項目値番号が残されているときにはステップ194に戻り、さもなければ、集計処理を終了する。このとき、2次元のクロス表が完成している。

- 上記本発明の第7実施例では、図1の表形式データのレコード全体を対象として集計を行っているが、レコードの部分集合を対象として同様の集計、例え  
20 ば、年齢16歳の人の性別／職業別の存在数を集計することが可能である。そのため、最初に、年齢16歳を検索条件として単一項目検索を行い、年齢16歳に該当するレコードの識別情報を結果集合として取得、保持する。次に、既に、複数項目AND検索に関して説明しているように、この結果集合に含まれるレコード、即ち、レコードの部分集合を対象として、結果集合に含まれるレ  
25 コードの識別情報を先頭から順次取り出し、レコードの識別情報に対応した項目値番号を値管理テーブルへのポインタ配列から取り出し、得られた項目値番号を2次元配列の行と列の座標とする要素の値のカウントを「1」ずつ加算す

ればよい。すなわち、この処理では、図49に示すように、まず、図9に示すものと略同等の処理を実行して、レコードへのポインタが格納されたような結果集合が生成される（ステップ4901）。次いで、この結果集合を用いて、図20と略同等の処理が実行される。ここでは、まず、図20の処理と同様に、  
5 集計に利用される項目に関する第1及び第2の情報ブロックが保持され（ステップ4902）、次いで、初期的な2次元配列が生成される（ステップ4903）。

その後、結果集合においてポインタの格納位置を示す番号（以下、場合によっては「格納位置番号」と称する）が初期化される（ステップ4904）。  
10 CPU100は、プログラムにしたがって、各格納位置番号に関して、第1及び第2の情報ブロックにおける値管理テーブルへの対応するポインタ配列を取り出され、2次元配列中の要素P（q1、q2）を特定し（ステップ4905）、次いで、その要素P（q1、q2）をインクリメントする（ステップ4906）。このような処理が、全ての格納位置番号に関して実行されることにより（つまり、結果集合中の前項目値番号に関して処理が施されることにより）、検索等により得られる部分集合に関するクロス集計をなすことができる。

次に、本発明の第8実施例につき説明を加える。この実施例においては、項目の項目値が幾つかのカテゴリに分類されている状況において、項目値の分類毎に存在数をカウントすることにより、クロス集計を実現している。例えば、  
20 図8に示された「職業」に関する情報ブロックを参照するに、「職業」には、「学生」、「プログラマ」、「教員」及び「その他」の4種類の項目値が登録されていることが分かる。この項目値に基づく分類を、新たな基準、例えば、収入の有無に従って、「所得者」、「無所得者」及び「不明」の3タイプに再分類した場合を想定する。本例では、この状況において、「収入の有無」とい  
25 う新しいカテゴリを作成し、性別／収入の有無による存在数のクロス集計を作成する。

図21に示された「職業」に関する情報ブロックは、特に、「収入の有無」

に基づいて各項目値番号毎に分類番号が付与された値管理テーブルを含む。本例では、学生には分類番号「1」（無所得）が割り与えられ、プログラマ及び教員には、分類番号「0」（有所得）、その他には、分類番号「2」（不明）が割り当てられている。

- 5      本発明の第8実施例にかかるクロス集計は、原則として第7実施例にかかるクロス集計と略同様の処理手順を有するが、集計データを格納する2次元配列の要素を指定する座標として、性別に関する第1の情報ブロックの項目値番号と、職業に関する第2の情報ブロックの分類番号とを使用する点が相違する。

- 10      2次元配列の要素の座標として項目値番号又は分類番号の何れでも使用できるようにするため、本発明の第8実施例では、第1及び第2の情報ブロックについて、各値管理テーブルへのポイント配列に格納された夫々の項目値番号を順次取り出し、値管理テーブルへのポイント配列から取り出された項目値番号自体又はその項目値番号に対応して値管理テーブルに格納された分類番号の何れかにより、2次元配列の要素Pの座標を特定する。

- 15      上述した例では、第1の情報ブロックとして「性別」の情報ブロックが用いられ、第2の情報ブロックとして「職業」の情報ブロックが用いられる（図20のステップ190参照）。次の処理ステップ（ステップ191）では、「性別」の情報ブロックには、2つの項目値番号が含まれ、その一方、「職業」の情報ブロックでは、3つの分類番号が含まれるため、 $2 \times 3$ （2行3列）の初期  
20      的な2次元配列が生成される。

これに引き続く処理ステップでも、第1のブロックの項目値番号q1と、第2のブロックの分類番号q2とが取り出され、これにより2次元配列中の1つの要素P（q1，q2）が特定され、要素Pの値が増加されるようになっている（ステップ194、196参照）。

- 25      上記第7実施例及び第8実施例によるクロス集計は、特に、存在数を求める形の集計であるが、本発明は、複数項目別（たとえば、性別／職業別）に平均年齢を求めるようなクロス集計にも拡張され得ることに注意すべきである。第

9 実施例は、上述したタイプのクロス集計をなしたものである。

上記本発明の第7実施例の如く、2次元配列の要素を1ずつインクリメントすることにより実現される存在数を求めるタイプの集計方法は、平均年齢を求めるような存在数の総和以外の演算を要するタイプのクロス集計に拡張するため、  
5 図22の動作フローチャートに示されるように変形される。

即ち、本発明の第9実施例によれば、集計のために2個の2次元配列が用いられ、第1の2次元配列に関しては、上記第7実施例と同様に性別／職業別の存在数をカウントし、第2の2次元配列に関しては、性別／職業別の年齢の総和を計算する。詳細に説明すると、

10 まず、性別、職業及び年齢の3個の項目について、第1、第2および第3の情報ブロックを記憶装置に保持する（ステップ200）。

性別及び職業の項目値番号のそれぞれの総数「2」及び「4」に対応して、集計データを格納する2個の $2 \times 4$ （2行4列）の初期的な2次元配列を作成する（ステップ202）。

15 第1及び第2の情報ブロックの値管理テーブルへのポインタ配列の先頭から順に項目値番号 $q_1$ 、 $q_2$ 及び $q_3$ を取り出し、2次元配列の要素の座標（ $q_1$ ， $q_2$ ）を特定し（ステップ204）、特定された第1の2次元配列の要素 $P_1$ （ $q_1$ ， $q_2$ ）の値を1ずつ増加する（ステップ206）。

さらに、「年齢」に関する情報ブロックについて、項目値番号 $q_3$ に対応した項目値（即ち、年齢）を取得し（ステップ208）、特定された第2の2次元配列の要素 $P_2$ （ $q_1$ ， $q_2$ ）の値に取得された年齢を加算する（ステップ210）。

このような処理の後に、対象となる全てのレコードが処理されたか否かを検査し（ステップ212）、処理されていない場合にステップ204に戻り、処理されている場合に、2次元配列 $P_1$ と2次元配列 $P_2$ の各要素の間で演算、即ち、 $P_2(q_1, q_2) / P_1(q_1, q_2)$ を算出する（ステップ206）。  
25 これにより、性別／職業別の平均年齢が得られ、平均のクロス集計表が作成さ

れる。

図23(A)は、上記本発明の第7実施例で得られるクロス集計表の概念図である。このように、上記第7実施例では、全ての性別／職業の組合せに関する存在数が集計される。しかし、図23(B)に示される如く、性別／職業別  
5    の中で、例えば、特に、性別が女性で、職業が学生である人の存在数ののみを知りたい場合がある。本発明によれば、この場合の存在数は、「女性」AND「学生」という複数項目AND検索の結果集合のサイズを求めることにより獲得される。

同様に、上記本発明の第9実施例では、全ての性別／職業の組合せに関して、  
10    平均年齢のクロス集計表を求めているが、特に、性別が女性で、職業が学生である人の平均年齢だけを求めることが可能である。その場合、「女性」AND「学生」という複数項目AND検索の結果集合のサイズから存在数を求めると共に、その結果集合に含まれるレコードの識別情報によって指定されるレコードに属する年齢を加算して年齢の総和を求め、(年齢の総和)／(存在数)を  
15    計算することにより平均年齢のクロス集計表における特定のセルに関する所望の値(たとえば、平均年齢)を求めることができる。

次に、本発明の第10実施例につき説明を加える。図24はマルチアンサ形式の項目を表す図であり、図25は本発明の第10実施例によるマルチアンサ項目対応型の情報ブロックの説明図である。マルチアンサとは、例えば、「今、  
20    机の上にある文房具は何ですか?」という質問に対する回答を得た場合に「鉛筆、消しゴム」又は「紙、鉛筆」のように同じ人から複数の回答が得られる状況を意味する。即ち、マルチアンサの場合、一つのレコードが一つの項目に関して複数の項目値を指定することが生じ得る。図24は上記の質問に対し、100万人から得た回答をそのまま一覧表にしたものである。

25    このようなデータを処理するため、本発明の第10実施例によれば、図25に示される如く、情報ブロックの値管理テーブルへのポインタ配列は、前述の項目値番号自体の配列と異なり、配列中のポインタに、項目値番号毎に1ビット

トが割り当てられている。したがって、レコードがその項目値番号を指定しているか否かをビットのオン／オフ（すなわち2進値）によって示すことができる。これにより、一つのレコードについて、一つの項目に含まれる複数の項目値を指定することが可能になる。たとえば、図25において、ポインタ配列中のポインタ（ビットポインタ）は、4ビットであり、最上位ビットがオン（すなわち「1」）であるときに、「紙」という回答が含まれたことを意味し、2番目のビットがオンであるときには、「定規」という回答が含まれ、3番目のビットがオンであるときには、「消しゴム」という回答が含まれることを意味している。さらに、最下位ビットがオンであるときには、「鉛筆」という回答が含まれることを意味している。

レコード番号「0」に対応するポインタは「3」である。これは「 $2^1 + 2^0$ 」と考えることができる。したがって、このレコード番号に対応して、「鉛筆」および「消しゴム」という回答が含まれていたことが理解できる。また、レコード番号「1」およびレコード番号「2」に対応するポインタは、それぞれ、「4」、「10」であり、これらは、それぞれ、「 $2^2$ 」および「 $2^3 + 2^1$ 」と考えることができる。したがって、これらレコード番号に対応して「定規」という回答、並びに、「消しゴム」および「紙」という回答が、それぞれ含まれることを知ることができる。

この実施例によれば、ポインタ値の各ビットに意味を持たせて、複数の項目値番号を示すことができるようになっている。したがって、レコードが複数の項目値を持つ場合であっても、ポインタ値によりこれを表現することが可能となる。

なお、本発明によれば、情報ブロックの一部分の構成を変形するだけでマルチアンサに容易に対応することができる点が有利である。實際上、このような変形を加えた情報ブロックに、上記本発明の種々の実施例で採用された情報ブロックを置き換えることができる。

次に、本発明の第11実施例につき説明を加える。図26は、本発明の第1

1 実施例に従って、集計処理の際に生じるブランク、エラー値等の特殊値の取扱方法を説明する図である。同図に示すように、第11実施例においては、ブランクをカテゴリーの1つとみなしてクロス集計を実行している。実際のデータを扱う場合には、ブランクや、 $\text{Log}(-1)$  のような数学的なエラーが出現  
5 することがある。本発明によれば、このような特殊値（ブランク、エラー等）であっても、値管理テーブルに項目値として登録し、登録された特殊値をそのまま検索や集計のカテゴリーとして使用できる点が有利である。

次に、本発明の第12実施例につき説明を加える。図27に示された本発明の第12実施例による複数項目検索方法の動作フローチャートを参照して遅延  
10 評価について説明する。本実施例では、本発明の第2実施例と同様に、図1に示した表形式データに関して「年齢」が「16歳」又は「19歳」であるという第1の検索条件と、「職業」が「学生」であるという第2の検索条件とを満たすレコードの集合を得る場合を考える。

上記本発明の第2実施例では、第2の検索条件に関して、全てのレコードの  
15 分類番号を予め設定しているが（図12のステップ124）、第12実施例の場合に、分類番号の設定は、第1の検索条件による検索の結果集合に基づいて実際にアクセスされた項目値番号に対応した分類番号だけに対して行われる。

既に説明したように、第1の項目である「年齢」に関する情報ブロックである第1の特定情報ブロックについて、「年齢」が「16歳」又は「19歳」であるレコードの結果集合は本発明の第1実施例に従って得られる（ステップ2  
20 20）。

次に、図8に示されている第2の項目である「職業」に関する情報ブロックである第2の特定情報ブロックとして選択し（ステップ222）、第2の特定情報ブロックの値管理テーブルにおける全ての分類番号の値を例えば、「  
25 1」に初期化する（ステップ224）。

次に、第1の検索条件に関する結果集合から、レコードへのポインタを表すレコード番号を順次取り出す（ステップ226）。本例の場合、図14に示さ



れるように、例えば、レコード番号「999999」が取り出される。

次に、第2の特定情報ブロックに関して、値管理テーブルへのポインタ配列から上記第1の検索条件について得られたレコード番号に対応した項目値番号を取り出す（ステップ228）。本例の場合、図14に示されるように、例え  
5 ば、レコード番号「999999」に対応した項目値番号「0」が取り出される。

次に、第2の特定情報ブロックについて取り出された項目値番号に対応する分類番号の値が「-1」であるか否かを検査する（ステップ230）。

分類番号が「-1」の場合には、当該項目値番号に対して未だ分類番号が設  
10 定されていないので、この項目値番号に対応した項目値が上記第2の検索条件に合致するかどうかを判定し（ステップ232）、合致する場合には、分類番号として「1」を設定し（ステップ234）、合致しない場合には、分類番号として「0」を設定する（ステップ236）。

分類番号の値が「-1」ではない場合、上記取り出された項目値番号に対応  
15 する分類番号の値が「1」に設定されているかどうかを判定する（ステップ238）。分類番号の値が「1」に設定されている場合、値管理テーブルへのポインタ配列内で、分類番号が「1」にセットされている当該項目値番号を指すポインタが格納されている場所に対応したレコードへのポインタ、例えば、レコード番号を最終的な結果集合に追加する（ステップ240）。本例では、図  
20 14に示されているように、例えば、レコード番号「999999」が最終的な結果集合に追加される。分類番号が「0」である場合には、最終的な結果集合は更新されない。

上記第12実施例のような遅延評価は、以下のような場合に有効である。たとえば、100万人の顧客名簿が存在し、実態調査のためのヒヤリングを実現  
25 したいが、都合により、100人のサンプルを抽出する場合を考える。たとえば、所定の条件（性別、年齢、職業など）を満たす人を絞り込んだところ、1万人にすることができ、さらに、無作為性を確保するために電話番号の末尾を

ある数（たとえば、「12」）として検索を行われるとする。

この場合に、第12実施例においては、まず、「分類番号配列」の要素を「-1」で埋めて、上記1万人の集合のみを評価する。つまり、1万人というサイズの結果集合について、その分類番号配列の要素を参照し、それが「-1」であれば、そのとき初めて電話番号を参照するように構成し、参照結果が「分類番号配列」の要素として与えられる。これにより、チェックの回数を1万回に抑えることが可能となる。このようにして、第12実施例によれば、通常のAND検索と比較して、処理ステップ数を大幅に減少させることが可能となる。

10      また、本発明にかかる情報ブロックを用いれば、電話番号のように「国番号+市外局番+局番+番号」のような構造を持つデータを複数の情報ブロックに分割して登録することにより、国番号、市外局番等の一部のデータに関する検索・集計が容易に行える利点が得られる。

15      また、既に説明した本発明の実施例による分類番号を利用することにより、例えば、年齢に関して、「10～19歳」を「10代」とし、「20～29歳」を「20代」とし、以下同様に取りまとめて新たなカテゴリー区分を生成することができ、かつ、生成された新しいカテゴリーに関して、前述と同様の検索・集計方法を適用できる。

20      前述したように、本発明の実施例にかかる検索・集計方法を実現する装置は、図28に示す一般的なコンピュータシステム、例えば、CPU100と、ROM110と、RAM120と、ハードディスク130と、ディスプレイ140のような出力装置と、KB/マウス150等の入力装置とがバス160を介して接続されているパーソナルコンピュータ上に実装することが可能である。したがって、前述したように、上記実施例を実現するための情報ブロックの構築  
25      するためのプログラム（情報ブロック生成プログラム）も、CD-ROMやROM110、ハードディスク記憶装置130に記憶されてもよいし、或いは、ネットワーク（図示せず）を介して外部から供給されてもよい。

図29に示されたフローチャートを参照して、一例として図2(B)に示された表形式データに対し、図5に示された形式の情報ブロックを構築する方法を以下に説明する。

#### ステップ300：データの準備

- 5      まず、図2(B)に示す形態のデータを用意する。次に、これを項目毎に分解する。図2(B)では、「性別」、「年齢」及び「職業」の項目に分解することができる。

#### ステップ311：「性別」項目用の情報ブロックの生成

- 10      1つの情報ブロックを生成し、これを、例えば、「性別」項目の情報ブロックとする。

#### ステップ312：値管理テーブルの作成

- 15      次に、値管理テーブルを初期化し、「性別」項目データを先頭から末尾まで走査することで、どのような項目が何件あるかを数えて格納する。本例の場合、ここまでのステップで、上記値管理テーブルの項目値配列11には、項目値として「女性」及び「男性」がセットされ、上記値管理テーブルの存在数配列14には、上記項目値に対応して、それぞれ、「367436」および「632564」がセットされることになる。

- 20      次に、項目値配列11中の項目値（「女性」及び「男性」）を所望の基準に従ってソートする。もちろん、そのソートの際には、存在数配列14も、項目値配列11のソートに伴って、並び替えられなければならない。

さらに、値管理テーブルの開始位置配列13中の開始位置をセットする。この開始位置は、値管理テーブル内の存在数配列14における最初の存在数から、該開始位置に対応する存在数の累計として求められる。もちろん、先頭の開始位置の値は、「0」である。

- 25      次に、開始位置配列13の内容を、分類番号配列12にコピーする。分類番号配列12は、後程、レコードへのポインタ配列を作成する際の作業領域として使用する。

ステップ3 1 3：値管理テーブルへのポインタ配列の作成

次に、値管理テーブルへのポインタ配列2 0のための領域を確保する。（領域のサイズは、上記存在数配列1 4における存在数の総和である。）

さらに、「性別」項目データを先頭から末尾まで、項目値を1つずつ取り出し、  
5 し、各項目値が、値管理テーブルの何番目の項目値と一致するかを調べ、  
「n」番目の項目値であった場合、「n-1」を値管理テーブルへのポインタ  
として上記値管理テーブルへのポインタ配列に格納する。

ステップ3 1 4：レコードへのポインタ配列の作成

次に、レコードへのポインタ配列3 0のための領域を確保する。本例では、  
10 領域のサイズは、上記存在数配列1 4における存在数の総和である。値管理テ  
ーブルへのポインタ配列2 0において、先頭の行から末尾の行まで、1つずつ  
値管理テーブルへのポインタを取り出す。値管理テーブルへのポインタ配列2  
0の「J」番目の値を取り出し、その値が「K」であり、値管理テーブルの  
「K+1」番目のレコードに対応する分類番号を取り出し、その値が「L」で  
15 あったとすると、レコードへのポインタ配列3 0の「L+1」番目の要素に  
「J-1」を格納し、値管理テーブルの「K+1」番目のレコードに対応する  
分類番号を1だけ増加させる。

以上の操作で、「性別」項目の情報ブロックの作成が完了した（ステップ3  
1 0）。「年齢」項目、「職業」項目も同様に情報ブロックを作成することが  
20 でき（ステップ3 2 0及びステップ3 3 0）、これにより、表形式データの全  
体に対する情報ブロックが得られる。

図3 0ないし図3 5は、図1に示された表形式データの「職業」に関する情  
報ブロックを作成する手順の説明図である。

図3 0には、属性値の種類が予め分かっているカテゴリー定義済みの場合の  
25 新規データ投入を示す図である。ここでは、既知のカテゴリー定義に従って、  
値管理テーブルが作成される。開始位置、存在数は不明であるため、「0」に  
初期化される。また、値管理テーブルへのポインタ配列、レコードへのポイン

タ配列の為の領域を確保し、これらも同様に初期化しておく。

図3 1には、値管理テーブルへのポインタ配列と、値管理テーブルの存在数を完成させるパスが示されている。投入したいデータを先頭からひとつずつ取りだし、それが値管理テーブルの何番目の値であるか（即ち、項目値番号）を調べ、値管理テーブルへのポインタ配列に格納し、値管理テーブル中の対応する存在数を「+1」ずつ更新する。図3 1の例は、投入データの上から2番目までの処理が終わった状態を示している。

図3 2には、値管理テーブルを完成させる第2のパスが示されている。存在数の累計は開始位置に対応することを利用して、開始位置が求められる。さらに、分類番号には開始位置の値がコピーされる。同図では分類番号の設定が終了している。

図3 3ないし図3 5には、データ投入の第3のパスが示されている。このパスは、値管理テーブルへのポインタ配列の先頭から1つずつ値（ポインタ）を取りだし、その値が参照する値管理テーブル内の分類番号により特定されるレコードへのポインタ配列中の位置に、値管理テーブルへのポインタ配列のオフセット、即ち、レコード番号を格納する。図3 3、3 4及び3 5には、夫々、「職業」に関する情報ブロックの値管理テーブルへのポインタ配列の先頭、2番目及び最後のデータについての処理が示されている。

なお、上記説明では、分類番号のフィールドが作業領域として使用されているが値管理テーブルの行数、即ち、項目値番号の総数と同数以上の要素を持つ整数配列であればいかなる配列でも作業領域として使用することができる。

一方、予めカテゴリー定義がなされていない場合の新規投入は、投入データを走査して値管理テーブルに登録すべき値の一覧を獲得した後、上記のカテゴリー定義がある場合の新規投入の処理を行うことにより実現される。

次に、図3 5に示されるような新規投入が完了した後の「職業」に関する情報ブロックに対し、「学生」という別のデータを追加する場合を考える。図3 6はかかるデータ追加を説明する図である。

この場合、最初に、値管理テーブルへのポインタ配列の末尾に「学生」を表す項目値番号0を追加し、値管理テーブルの中の学生の存在数を「1」だけ増やす。次に、値管理テーブルへのポインタ値(=1000000)、即ち、レコード番号を格納するための領域をレコードへのポインタ配列内に確保する必要がある。そのため、レコードへのポインタ配列の中の「学生」に対応した領域の末尾の値(本例では、999999)を取り出し、そこに、拡張先のアドレスである「1000000」を格納する。但し、拡張先のアドレスであることが識別できるように、例えば、符号を反転させて、「-1000000」として格納する。続いて、拡張先に先に取り出された末尾の値である「999999」を格納し、最後に、新たに追加されたデータに対応したポインタ値「1000000」を格納する。

このような追加方法を採用すると、データを追加する際に大量のデータを移動させる必要が回避される。また、多数のデータ追加が行われた場合に、ポインタの数が増加することに起因したアクセスの効率低下を抑えるために、適当なタイミングで新規投入の第3のパスと同じ処理を繰り返しても良い。

図37には、本発明の他の実施例による情報ブロックの構造が示されている。同図に示された構造を採用した場合、ポインタの数の増加は回避され容易にデータ変更を行うことが可能である。図37において、開始位置配列には、レコードへのポインタ配列が配置されている領域の先頭を示すアドレスが格納されている。たとえば、「学生」という項目値に関しては、開始位置として「0」が格納されている。その一方、「プログラマ」という項目値に関しては、開始位置として「n(ただし、 $n > 455214$ )」という値が割り当てられている。

次に、本発明の第13実施例につき説明を加える。この実施例では、上記情報ブロックを用いたレコードのソートを実現している。図38には「職業」という項目に関するレコードのソートの初期状態が示されている。同図に示された原データはソートされるべきレコード番号の配列を表す。例えば、「職業」

以外の項目について得られたレコードへのポインタ配列、或いは、検索の結果集合が原データとして利用され得る。本例の場合、説明の便宜上、原データのレコード番号は、“0，1，2，．．．，9”の順番に配列されているが、一般にソート前のレコード番号の並びは乱雑であることに注意する必要がある。

- 5 各レコード番号に対応した「職業」の項目値が“教員，プログラマ，学生，．．．，その他”のように順番に配列されている。

同図の右側には、「職業」に関する情報ブロックに含まれる種々の配列の初期状態が示されている。「職業」に関する情報ブロックは、例えば、図29ないし図35を参照して説明した情報ブロックの構築方法に従って作成されている。

10 「職業」に関するレコードのソートを行うとき、予め作成された「職業」に関する情報ブロックの値管理テーブル、及び、値管理テーブルへのポインタ配列が利用される。開始位置として、情報ブロックの構築の際に設定された開始位置がそのまま使用される。終了位置には対応する開始位置の値がコピーされている。終了位置を格納する領域は、例えば、存在数のために確保された領域（存在数配列）を用いても構わない。値管理テーブルへのポインタ配列は、

15 例えば、レコード番号順に予め作成されている。本例の場合、原データのレコード番号の配列がレコード番号の昇順であるため、原データのレコード番号と、値管理テーブルへのポインタ配列とのつながりは単純な関係で示されている。また、レコードへのポインタ配列は、ソートされた結果集合を格納するための

20 配列であり、ソートされるデータと同じサイズが確保される。上記終了位置は、レコードへのポインタ配列にソートされた結果を格納するためのアドレスとして使用される。

図39は、本発明の第13実施例によるソートの第1ステップの状態を説明する図である。第1ステップでは原データの先頭のレコード（本例では、レコード番号=0）が処理される。レコード番号「0」のレコードの項目「職業」に関する項目値は「教員」である。このとき、レコード番号「0」に対応した値管理テーブルへのポインタ配列には、項目値「教員」を指定する項目値番号

25

「2」が格納されている。そこで、項目値番号「2」に対応した終了位置の値「5」を取り出し、この値「5」をアドレスとして、ソートされた結果集合が格納されるレコードへのポインタ配列の5番目に、当該レコード番号「0」を設定する。次に、当該項目値番号「2」に対応した終了位置の値は「+1」され、

5 「5」から「6」に増加する。

図40は、本発明の第13実施例によるソートの第2ステップの状態を説明する図である。第2ステップでは原データの2番目のレコード（本例では、レコード番号=1）が処理される。レコード番号「1」のレコードの項目「職業」に関する項目値は「プログラマ」である。このとき、レコード番号「1」

10 に対応した値管理テーブルへのポインタ配列には、項目値「プログラマ」を指定する項目値番号「1」が格納されている。そこで、項目値番号「1」に対応した終了位置の値「3」を取り出し、この値「3」をアドレスとして、ソートされた結果集合が格納されるレコードへのポインタ配列の3番目に、当該レコード番号「1」を設定する。次に、当該項目値番号「1」に対応した終了位置

15 の値は「+1」され、「3」から「4」に増加する。

以下、原データの残りのレコード番号“2, 3, 4, 5, 6, 7, 8, 9”に対し、上記第1又は第2ステップと同じ動作を繰り返し適用する。かくして得られたソートの最終状態が図41に示されている。同図のソートされた結果集合から判るように、本発明の第13実施例によるソートの結果として、レ

20 コードは、「職業」に関する項目値番号の順にソートされ、レコード番号“2, 4, 6, 1, 7, 0, 3, 5, 8, 9”の順に並べ換えられている。

上記第13実施例の説明では、原データが元の表形式データの全てのレコードを含む場合、即ち、全体集合を想定している。しかし、本発明によるソートは、一部のレコード、即ち、部分集合に対しても有効である。以下、図42及び

25 図43を参照して部分集合に対するソートを説明する。

図42には、部分集合の全要素に対し、上記のソート処理が完了した状態を示す図である。本例の場合、原データは、「職業」についての項目値が「教



員」であるレコード番号「0」のレコードと、項目値が「プログラマ」であるレコード番号「1」のレコードが原データとして与えられている。これら両方のレコードに対しソート処理を適用すると、同図のソートされた結果集合に示されるようなソート結果が得られる。この時点では、結果集合はレコードへの  
5 ポインタ配列に格納されている。したがって、部分集合のソート結果を格納するため全体集合と同じサイズの領域が確保されている。

そこで、部分集合に対するソートの場合、結果集合を部分集合と同じサイズに圧縮することが好ましい。図43には、かかる部分集合に対するソートの後  
10 処理が示されている。この後処理、即ち、結果集合の圧縮は、値管理テーブルの各項目値に関する開始位置と終了位置の差から、当該項目値に対応したソート結果の存在数及び格納位置を取得し、取得された存在数及び格納位置に基づいてソート結果を取り出して並べる。

上記全体集合或いは部分集合に関するソート処理につき、再度、図50のフローチャートを利用して一般的な説明を加える。ここでは、全体集合および部  
15 分集合の何れの場合にも、原データには、先頭から順番に、格納位置番号が付与されていると考える。たとえば、図38および図41では、格納位置番号はレコード番号と一致する。しかしながら、原データの並び順が、レコード番号にしたがっていない場合には、格納位置番号とレコード番号とは相違する。なお、上記実施例に関して具体的に説明したソート処理等が、CPU100が所  
20 定の領域に格納されたプログラムを実行することにより実現できることは言うまでもない。

ソート処理においては、まず、格納位置番号が初期化される（ステップ5001）。次いで、ある格納位置番号に関して、値管理テーブルへのポインタ配列中の対応するポインタが参照され（ステップ5002）、当該ポインタの指  
25 定する項目値番号に位置する終了位置の値が特定される（ステップ5003）。

その後、上記終了位置にて特定される、レコードへのポインタ配列中の位置に、対応するレコード番号が格納される（ステップ5004）。その後、ス

テップ 5 0 0 3 において特定された終了位置の値がインクリメントされる(ステップ 5 0 0 5)。上記ステップ 5 0 0 2 ないしステップ 5 0 0 5 の処理は、全ての原データに関して施され(ステップ 5 0 0 6、5 0 0 7 参照)、これにより、所定のレコード番号が配置されたレコードへのポインタ配列を得ることができる。図 3 9 の例は、図 5 0 の処理のうち、格納位置番号が「0」である場合のステップ 5 0 0 2 ～ステップ 5 0 0 5 に対応することが理解でき、かつ、図 4 0 は、格納位置番号が「1」である場合のステップ 5 0 0 2 ～ステップ 5 0 0 5 に対応することが理解できるであろう。

また、部分集合に対するソート処理の場合には、ステップ 5 0 0 1 ～ステップ 5 0 0 7 が実行された後に、ソートの後処理(ステップ 5 0 0 8 参照)により、結果集合を圧縮すればよい。

上記本発明の第 1 3 実施例によるソートは、所謂「昇順」のソートであり、即ち、ソート結果は、ソートされる項目値の項目値番号が増加する順に並べられている。しかし、ソート結果は、「降順」、即ち、ソートされる項目値の項目値番号が減少する順に並べられることもできる。「降順」ソートは、「昇順」ソートの場合に使用された開始位置を修正することにより実現される。本例の場合、「昇順」ソートの開始位置は、

学生	0
プログラマ	3
20 教員	5
その他	6

である。これに対し、「降順」ソートの開始位置は、

その他	0
教員	$10 - 6 = 4$
25 プログラマ	$10 - 5 = 5$
学生	$10 - 3 = 7$

のようになる。

このような本発明の第13実施例にしたがったソートによれば、以下の利点  
が得られる。

第1に、高速ソートが実現される。例えば、Pentium Pro(登録商標)  
200MHz/Windows95(登録商標)という動作環境において、本発明による新規の  
5 ソートは、100万件当たりのソート時間が145 msecであった。その一  
方、従来のクイックソートの場合、100万個の整数をソートするため要する  
時間は1530 msecであった。

第2に、値として格納される項目値の種類にかかわらず一定の性能が得られ  
る。本ソートによれば、例えば、項目値のデータ型が文字列型若しくは浮動小  
10 数点型であっても、整数型の場合と変わらない性能を得ることができる。これ  
に対し、従来のクイックソートなどの方式の場合、取り扱うデータが整数型で  
ある場合に最も高速であり、可変長の文字列型である場合に最も遅くなる。

第3に、データサイズが増加してもソート速度が低下しない。本ソートでは、  
ソートの速度は、データサイズを $n$ として、 $O(n)$ で表される。一方、従来  
15 の、例えば、クイックソートのような高速ソート法では、ソートの速度は、 $O(n \cdot \log(n))$ である。

第4に、複数項目のソートを項目毎のソートに分割することができる。例え  
ば、図38ないし図41を参照して説明した本発明の第13実施例において、  
原データ中、同じ項目値「学生」に対応したレコードは、レコード番号「2」、  
20 レコード番号「4」及びレコード番号「6」の順に並べられている。そして、  
最終的なソートされた結果集合において、このレコード番号の並び順（即ち、  
レコード番号「2」、レコード番号「4」及びレコード番号「6」）が保存さ  
れている。これは、ソートされた結果のレコードの並び順は、ソート前のレコ  
ードの並び順をソートの目的を満たす範囲内で反映していることを意味する。  
25 このような本ソートの特徴を利用することにより、項目毎のソートを順番に行  
うことにより、複数の項目のソートが実現される。これに対し、従来のクイッ  
クソートでは、ソート結果の並び順にソート前の状態は反映されないことが知

られている。

### 具体例

上記本発明の各種実施例において、値管理テーブルには、項目値に関する値リストが与えられている。例えば、図7に示された例では、項目値の列には、

5 「16歳」、「17歳」、「18歳」、... という値を有するリストが存在する。また、値管理テーブルには、項目値番号毎に分類番号が設けられている。以下、このような項目値に関する値リストと分類番号との組合せが、ある項目に関する多数の値に対して、少数回の比較判断によって検索条件の可否を直ちに決定できることを示す具体例を、以下に説明する。

- 10 例えば、値リストとして、値の大きさに関してソートされた以下のリスト：
- 0.1, 0.2, ...,
- 100.0, 100.1, 100.2, ...,
- 1000.0

- 15 が与えられ、検索条件として、例えば、「値が100より大」が与えられた場合を想定する。

- 最初に、条件を満たさない最小の値（本例では、100）を値リストから探す。そして、最小の値、即ち、「100」よりも前にある値リスト中の値に対応した分類番号をすべて「0」にセットする。また、「100」よりも後にある値リスト中の値に対応した分類番号をすべて1にセットする。これにより、
- 20 最小の値を探し出すことができれば、それ以降、いっさいの比較演算を行うことなく分類番号がセットされ、検索条件を満たす値を有する項目値若しくは項目値番号が得られる。

- 最小の値は、従来技術において公知である2分法等の方法を用いることにより少数回の比較演算で見つけることができる。例えば、値リスト中に存在する
- 25 値のバリエーションがN個であるとする、2分法によって上記の最小の値を見つけるために要する比較演算の回数は、 $\log_2(N)$  回程度である。

一方、値リストが値の大きさに関してソートされていない場合、同様に値リ

スト中に存在する値のバリエーションがN個あるとき、検索条件を満たす値を見つけるためにはN回の比較演算が必要とされる。

このように、値管理テーブルにおいて、値リストがソートされ、かつ、分類番号が設けられることにより、所定の検索条件の可否の決定が高速化される。

- 5      また、当業者には当然に理解されるように、上記の値リストの内容、並びに、検索条件は本例を説明するための具体的な一例に過ぎず、本発明によれば、種々の値リスト、並びに、種々の検索条件の組合せについても同様に検索条件の可否の決定が高速化される。

- 次に、100万件のデータの検索・集計のテストについて説明する。テスト  
10      に使用したプラットフォームは、Pentium Pro（登録商標）200MHz と、128MB メモリを搭載した一般的なパーソナルコンピュータである。図44はテストに使用したデータの一覧を示す図表である。「000000」～「999999」までの100万個の数字を、10000の単位、100の単位、1の単位の3項目に分解した表形式データである。各項目に「00」～「99」までの  
15      項目値が1万回ずつ出現する。

図45は、100万件の検索／集計に要する時間を使用した結果集合のタイプ毎に測定した測定結果の一覧である。結果集合のタイプは、上記のビットフラグ形式とポインタ配列形式の2種類である。測定結果の時間はミリ秒（1／1000秒）単位である。

- 20      上記テストで行った検索は、「×1万」、「×100」、「×1」の3項目をAND条件で接続した複数項目AND検索である。検索は、「×1万」、「×100」、「×1」の項目順にカスケードした。検索の中間及び最終の結果集合は、上記の通りビットフラグ又はポインタ配列のいずれかの形式である。測定時間は5回の測定の平均である。

- 25      また、上記テストの集計は、検索のテストにより得られた結果集合に対して、「×100」、「×1」の各値（00～99）が出現した回数の計数である。このクロス集計の表のサイズは、（100×100＝）1万セルである。時間

は全てミリ秒単位であり、5回の測定の平均を測定時間とした。

また、表形式データの検索・集計システムの構成は、上記の実施例で説明された例に限定されることなく、検索・集計システムの各々の構成要件をソフトウェア（プログラム）で構築し、ディスク装置等に格納しておき、必要に応じて検索・集計システムのコンピュータにインストールして表形式データの検索・集計を行うことも可能である。さらに、構築されたプログラムをフロッピーディスクやCD-ROM等の可搬記憶媒体に格納し、このようなシステムを用いる場面で汎用的に使用することも可能である。

本発明は、以上の実施の形態に限定されることなく、特許請求の範囲に記載された発明の範囲内で、種々の変更が可能であり、それらも本発明の範囲内に包含されるものであることは言うまでもない。

たとえば、上記第2実施例に示すAND検索およびOR検索においては、2種の項目に関する検索を実行しているが、これに限定されるものではなく、3以上の複数種の項目に関する検索も実現できることは明らかである。

また、上記第7実施例および第8実施例においては、2種の項目に関する集計のために、2次元配列を生成しているが、これに限定されるものではなく、3種以上の項目に関する集計のために、3次元以上の多次元配列を生成し、これを利用して上記集計が可能であることはいうまでもない。3種の項目に関する集計を考えると、3つの情報ブロックの各々における項目値番号 $q_1$ 、 $q_2$ 、 $q_3$ を取り出し、これにより、3次元配列の一個の要素 $P(q_1, q_2, q_3)$ が特定される。

さらに、上記第9実施例に関しても、第7実施例および第8実施例と同様に、3種以上の項目に関する集計をなすことができることは言うまでもない。

また、上記実施の形態においては、一般のコンピュータシステム内に、所定のプログラムを読み込み、当該プログラムを実行することにより、検索処理、集計処理および／またはソート処理を実現しているが、本発明はこれに限定されるものではなく、パーソナルコンピュータ等のような一般のコンピュータシステムに、

データ処理専用のボードコンピュータを接続し、当該ボードコンピュータが上記処理を実行できるように構成しても良いことは言うまでもない。したがって、本明細書において、手段とは必ずしも物理的手段を意味するものではなく、各手段の機能が、ソフトウェアによって実現される場合或いはその一部または全部がハードウェアにより実現される場合を包含する。さらに、一つの手段の機能が、二つ以上の物理的手段により実現されても、若しくは、二つ以上の手段の機能が、一つの物理的手段により実現されてもよい。上記の説明の通り、本発明によれば、表形式で表現される大量のデータを、アクセスに時間がかかる従来のデータ表を用いることなく処理することができるので、集計・検索のスピードを大幅に向上させることができる。

#### 産業上の利用分野

本発明は、特に、大量のデータを管理するシステム、たとえば、データベース、データウェアハウスに利用することができる。より具体的には、大規模な科学技術計算、プラントや電力供給などの制御システム、配送や資源分配等の計画法、受発注管理や証券取引などの事務管理に利用可能である。

## 請求の範囲

1. 情報の項目に対応した項目値を含むレコードの配列として表される表形式データから、特定の項目及び特定のレコードに対応した項目値を得る方法であって、

5       上記特定の項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とを記憶装置に保持し、  
      上記項目値番号指定情報配列から上記特定のレコードに対応した項目値番号を取得し、

10       上記値管理テーブルに格納された項目値の中から、上記取得された項目値番号に対応した項目値を得ることを特徴とする方法。

2. 上記項目値番号に対応した項目値を分類するための分類番号が上記項目値番号に対応して上記値管理テーブルに格納され、

      上記項目値番号に対応した項目値を得る際に上記分類番号が参照されることを特徴とする請求の範囲第1項に記載の方法。  
15

3. 情報の項目に対応した項目値を含むレコードの配列として表される表形式データから検索条件に適合する項目値を含むレコードを検索する方法であって、

      上記検索条件と関連した項目に属する項目値に対応した項目値番号の順番  
20       に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列と、相等しい項目値番号に関連する一以上のレコード識別情報を、上記項目値番号毎に固有の領域に格納したレコード識別情報配列とを含む個別項目情報であって、値管理テーブルが、上記項目値番号ごとに、上記レコード識別情報配列  
25       において相等しい項目値番号に関連する上記一以上のレコード識別情報を格納した領域を示すレコード識別情報指定情報を含むような、個別項目情報を記憶装置に保持し、



上記値管理テーブルに格納された項目値の中で、上記検索条件に適合する項目値と関連した項目値番号に対応する上記レコード識別情報指定情報を用いて、上記レコード識別情報配列から上記検索条件に適合するレコード識別情報を取り出すことを特徴とする方法。

- 5     4. 上記検索条件に適合するレコードの識別情報に基づいて上記検索条件に適合するレコードを指定する結果集合を記憶装置に保持し、

他の検索条件と関連した項目に関する他の個別項目情報に関して、

上記他の個別項目情報の項目値番号指定情報配列から、上記結果集合で指定された上記検索条件に適合したレコードの識別情報に対応した項目値番号  
10     を取り出し、

上記他の個別項目情報について上記取り出された項目値番号と関連した項目値が上記他の検索条件と適合するか否かを判定し、

上記他の個別項目情報について上記項目値番号と関連した項目値が上記他の検索条件と適合する場合に、上記項目値番号に対応した上記レコード識別  
15     情報を、上記別の検索条件に適合するレコード識別情報として取り出し、

必要な場合に、さらに他の検索条件に関して、さらに他の個別項目情報について、上記項目値番号の取り出し、適合判定およびレコード識別情報の取り出しを繰り返すことを特徴とする請求の範囲第3項に記載の方法。

5. 上記検索条件に適合するレコードの識別情報に基づいて上記検索条件に適合するレコードを指定する結果集合を記憶装置に保持し、  
20

他の検索条件と関連した項目に関する他の個別項目情報を記憶装置に保持し、

- 25     他の個別項目情報の値管理テーブルに格納された項目値の中で、上記検索条件に適合する項目値と関連した項目値に対応するレコード識別情報指定情報を用いて、レコード識別情報配列から、上記他の検索条件に適合するレコード識別情報を取り出し、検索条件に適合するレコードを指定する他の結果集合を記憶し、

必要な場合に、さらに他の検索条件に関して、さらに他の個別項目情報について、上記個別項目情報の記憶、レコード識別情報の取り出し、および、さらに他の結果集合の記憶を繰り返して、

5 得られた結果集合から重複したレコードを除去した最終的な結果集合を得ることを特徴とする請求の範囲第3項に記載の方法。

6. 上記値管理テーブルが、上記項目値番号ごとに、上記固有の領域の始点を示す開始位置と、相等しい項目値番号を備えたレコード数を示す存在数とを含み、

10 上記開始位置および存在数を参照して、所定のレコード識別情報配列を特定することを特徴とする請求の範囲第3項ないし第5項の何れか一項に記載の方法。

7. 上記値管理テーブルに、上記項目値番号に対応した項目値を分類するための分類番号が、上記項目値番号に対応して格納され、

15 上記分類番号を用いて、検索条件に適合する項目値を特定することを特徴とする請求の範囲第3項ないし第6項の何れか一項に記載の方法。

8. 情報の項目に対応した項目値を含むレコードの配列として表される表形式データを集計する方法であって、

20  $n$ が1以上の整数を表す場合に、集計に利用される $n$ 個の項目について、上記項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とにより表された $n$ 個の個別項目情報を記憶装置に保持し、

25  $i$ が $1 \leq i \leq n$ なる整数を表すとき、 $i$ 番目の個別情報項目について、上記項目値番号の総数が $N_i$ で表され、 $k_i$ が $0 \leq k_i \leq N_i - 1$ なる整数を表し、 $M$ が1以上の整数を表し、 $m$ が $1 \leq m \leq M$ なる整数であるとき、 $N_1 \times N_2 \times \dots \times N_i \times \dots \times N_n$ のサイズを持つ $n$ 次元の $M$ 個のデータ空間の各要素 $P_m(k_1, k_2, \dots, k_i, \dots, k_n)$ を初期化し、

上記  $n$  個の個別情報項目について、 $j$  が  $0 \leq j \leq (\text{レコードの総数} - 1)$  なる整数を表すとき、各項目値番号指定情報配列の  $j$  番目に格納された夫々の項目値番号を取り出し、 $i$  番目の個別情報項目から取り出された項目値番号が  $q_i$  で表されるとき、上記データ空間の要素  $P_m(q_1, q_2, \dots, q_i, \dots, q_n)$  を特定し、

上記特定された要素  $P_m(q_1, q_2, \dots, q_i, \dots, q_n)$  の値を処理するステップを含むことを特徴とする方法。

9. 情報の項目に対応した項目値を含むレコードの配列として表される表形式データを集計する方法であって、

$n$  が 1 以上の整数を表し、集計に利用される  $n$  個の項目について、上記項目に属する項目値に対応した項目値番号の順番に上記項目値及び上記項目値の分類番号が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とにより表された  $n$  個の個別項目情報を記憶装置に保持し、

$i$  が  $1 \leq i \leq n$  なる整数を表すとき、 $i$  番目の個別情報項目について、上記項目値番号又は上記分類番号の何れか一方の番号の総数が  $N_i$  で表され、 $k_i$  が  $0 \leq k_i \leq N_i - 1$  なる整数を表し、 $M$  が 1 以上の整数を表し、 $m$  が  $1 \leq m \leq M$  なる整数であるとき、 $N_1 \times N_2 \times \dots \times N_i \times \dots \times N_n$  のサイズを持つ  $n$  次元の  $M$  個のデータ空間の各要素  $P_m(k_1, k_2, \dots, k_i, \dots, k_n)$  を初期化し、

上記  $n$  個の個別情報項目について、 $j$  が  $0 \leq j \leq (\text{レコード総数} - 1)$  なる整数を表すとき、各項目値番号指定情報配列の  $j$  番目に格納された夫々の項目値番号を取り出し、 $i$  番目の個別情報項目から取り出された項目値番号又は上記  $i$  番目の個別情報項目の値管理テーブルに該項目値番号と対応して格納された分類番号が  $q_i$  で表されるとき、上記データ空間の要素  $P_m(q_1, q_2, \dots, q_i, \dots, q_n)$  を特定し、

上記特定された要素  $P_m(q_1, q_2, \dots, q_i, \dots, q_n)$  の値を

処理するステップを含むことを特徴とする方法。

10.  $M = 1$  であり、かつ、上記特定された要素  $P_m$  の値を処理するステップが、  
上記要素  $P_m$  の現在の値に 1 を加えることを特徴とする請求の範囲第 9 項または第 10 項に記載の方法。

5 11. 上記特定された要素  $P_m$  の値を処理するステップが、  
 $M$  個の要素  $P_m$  の中の少なくとも 1 個の要素  $P_m$  に対して、  
記憶装置に保持された別の個別項目情報について項目値番号指定情報配列の  $j$  番目に格納された項目値番号を取得し、

上記別の個別項目情報の値管理テーブルに格納された項目値の中から、上  
10 記取得された項目値番号に対応した項目値を取得し、  
上記要素  $P_m$  の現在の値と上記取得された項目値を組み合わせることで上記要素  $P_m$  の値を更新することを特徴とする請求の範囲第 9 項または第 10 項に記載の方法。

12. 上記項目値番号を指定する情報が、項目値番号自体であることを特徴とする  
15 請求の範囲第 1 項ないし第 11 項の何れか一項に記載の方法。

13. 上記項目値番号を指定する情報が、各項目値番号毎に 1 ビット以上が割り  
当てられ、指定の有無が設定される 2 進値であることを特徴とする請求の範囲第 1 項ないし第 11 項の何れか一項に記載の方法。

14. 情報の項目に対応した項目値を含むレコードの配列として表される表形式  
20 データを検索・集計する装置であって、

上記項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とを保持する記憶装置と、

上記記憶装置に保持された上記項目値番号指定情報配列から上記特定のレ  
25 コードに対応した項目値番号を取得する項目値番号取得手段と、

上記記憶装置に保持された上記値管理テーブルに格納された項目値の中から、上記取得された項目値番号に対応した項目値を得る項目値取得手段とを

有することを特徴とする装置。

15. 上記記憶装置が、上記値管理テーブルと、上記項目値番号指定情報配列と、  
相等しい項目値番号に関連する一以上のレコード識別情報を、上記項目値番  
号ごとに固有の領域に格納したレコード識別情報配列とを有する個別項目情  
5 報を保持し、かつ、上記値管理テーブルが、上記項目値番号ごとに、上記レ  
コード識別情報配列において相等しい項目値番号に関連する上記一以上のレ  
コード識別情報を格納した領域を示すレコード識別情報指定情報を含み、

さらに、上記値管理テーブルに格納された項目値の中で、所定の検索条件  
に適合する項目値と関連した項目値番号に対応する上記レコード識別情報指  
10 定情報を用いて、上記レコード識別情報配列から上記検索条件に適合するレ  
コード識別情報を取得する検索手段を有することを特徴とする請求の範囲第  
11項に記載の装置。

16. 情報の項目に対応した項目値を含むレコードの配列として表される表形式  
データを検索・集計するプログラムを記録したコンピュータにより読み出し  
15 可能な記憶媒体であって、

上記項目に属する項目値に対応した項目値番号の順番に上記項目値が格納  
されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定  
する情報が格納された項目値番号指定情報配列とを記憶装置に保持するステ  
ップと、

20 上記記憶装置に保持された上記項目値番号指定情報配列から上記特定のレ  
コードに対応した項目値番号を取得するステップと、

上記記憶装置に保持された上記値管理テーブルに格納された項目値の中か  
ら、上記取得された項目値番号に対応した項目値を得るステップとを有する  
ことを特徴とするプログラムを記録した記憶媒体。

25 17. 情報の項目に対応した項目値を含むレコードの配列として表される表形式  
データから検索条件に適合する項目値を含むレコードを検索するプログラム  
を記憶したコンピュータにより読み出し可能な記憶媒体であって、上記プロ

グラムが、

上記検索条件と関連した項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列と、相等しい項目値番号に関連する一以上のレコード識別情報を、上記項目値番号毎に固有の領域に格納したレコード識別情報配列とを含む個別項目情報であって、値管理テーブルが、上記項目値番号ごとに、上記レコード識別情報配列において相等しい項目値番号に関連する上記一以上のレコード識別情報を格納した領域を示すレコード識別情報指定情報を含むような、個別項目情報を記憶装置に保持するステップと、

上記値管理テーブルに格納された項目値の中で、上記検索条件に適合する項目値と関連した項目値番号に対応する上記レコード識別情報指定情報を用いて、上記レコード識別情報配列から上記検索条件に適合するレコード識別情報を取り出すステップとを有することを特徴とする記憶媒体。

18. 請求の範囲第1項ないし第13項の何れか一項に記載の方法を実現する、コンピュータのメモリ中にロード可能なコンピュータプログラムプロダクト。
19. 情報の項目に対応した項目値を含むレコードを指定するレコード識別情報の配列を上記項目に関して並べ換えるレコードのソート方法であって、

上記レコード毎に、上記レコード識別情報と上記項目の項目値に対応した項目値番号とを関連付け、

上記項目値番号毎に、上記レコード識別情報の並び換え後の格納場所を定義し、

上記配列から上記レコード識別情報を順番に取り出し、

上記取り出されたレコード識別情報に対応した上記項目値番号を判定し、

上記取り出されたレコード識別情報を、上記判定された項目値番号に対応して定義された上記格納場所に格納し、

上記レコード識別情報が格納される上記格納場所を、次のレコード識別情

報の格納のために更新することを特徴とする方法。

20. 上記検索条件と関連した項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とを含む個別項目情報を記憶装置に保持し、値管理テーブルが、さらに、上記項目値番号ごとに、上記レコード識別情報配列において相等しい項目値番号に関連する上記一以上のレコード識別情報を格納した領域を示すレコード識別情報指定情報を含み、

上記格納するステップが、上記レコード識別情報指定情報にしたがって、レコード識別情報を、上記格納場所に格納するように構成されたことを特徴とする請求の範囲第19項に記載の方法。

21. 上記レコード識別情報指定情報が、上記格納場所の領域の始点を示す開始位置と、初期的には開始位置と等価であり、上記更新によりその値がインクリメントされる終了位置とを含むことを特徴とする請求の範囲第20項に記載の方法。

22. 情報の項目に対応した項目値を含むレコードを指定するレコード識別情報の配列を上記項目に関して並べ換えるレコードのソート装置であって、

上記レコード毎に、上記レコード識別情報と上記項目の項目値に対応した項目値番号とを関連付ける手段と、

- 上記項目値番号毎に、上記レコード識別情報の並び換え後の格納場所を定義する定義手段と、

上記配列から上記レコード識別情報を順番に取り出して、取り出されたレコード識別情報に対応した上記項目値番号を判定する項目値番号判定手段と、

- 上記取り出されたレコード識別情報を、上記判定された項目値番号に対応して定義された上記格納場所に格納するレコード識別情報格納手段と、

上記レコード識別情報が格納される上記格納場所を、次のレコード識別情報の格納のために更新する更新手段とを備えたことを特徴とする装置。

23. 上記検索条件と関連した項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とを含む個別項目情報とを保持する記憶装置を備え、

5 値管理テーブルが、さらに、上記項目値番号ごとに、上記レコード識別情報配列において相等しい項目値番号に関連する上記一以上のレコード識別情報を格納した領域を示すレコード識別情報指定情報を含み、

上記レコード識別情報格納手段が、上記レコード識別情報指定情報にしたがって、レコード識別情報を、上記格納場所に格納するように構成されたこと  
10 とを特徴とする請求の範囲第22項に記載の装置。

24. 情報の項目に対応した項目値を含むレコードを指定するレコード識別情報の配列を上記項目に関して並べ換えるレコードのソートプログラムを記憶したコンピュータにより読み出し可能な記憶媒体であって、上記ソートプログラムが、

15 上記レコード毎に、上記レコード識別情報と上記項目の項目値に対応した項目値番号とを関連付けるステップと、

上記項目値番号毎に、上記レコード識別情報の並び換え後の格納場所を定義するステップと、

上記配列から上記レコード識別情報を順番に取り出すステップと、

20 上記取り出されたレコード識別情報に対応した上記項目値番号を判定するステップと、

上記取り出されたレコード識別情報を、上記判定された項目値番号に対応して定義された上記格納場所に格納するステップと、

上記レコード識別情報が格納された上記格納場所を、次のレコード識別情報の格納のために更新するステップとを備えたことを特徴とする記憶媒体。  
25

25. 請求の範囲第19項ないし第21項の何れか一項に記載の方法を実現する、コンピュータのメモリ中にロード可能なコンピュータプログラムプロダクト。



図 1 (従来技術)

レコード番号	性別	年齢	職業
0	女性	1 8	プログラマ
1	男性	2 1	学生
2	女性	3 1	教員
9 9 9 9 9 9	女性	1 6	学生

図 2 (従来技術)

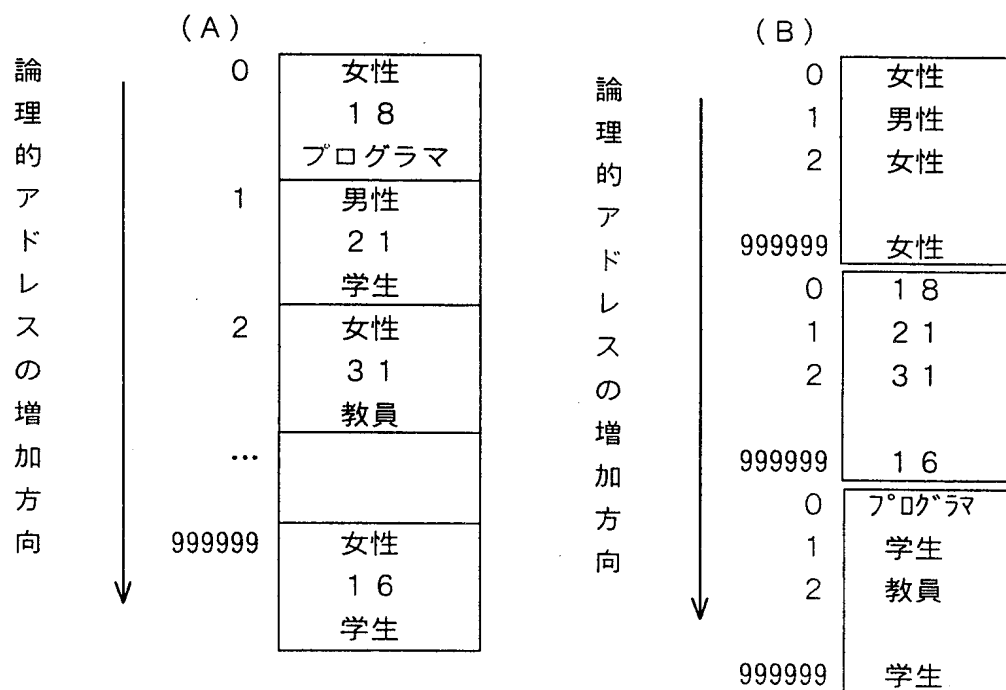


図28

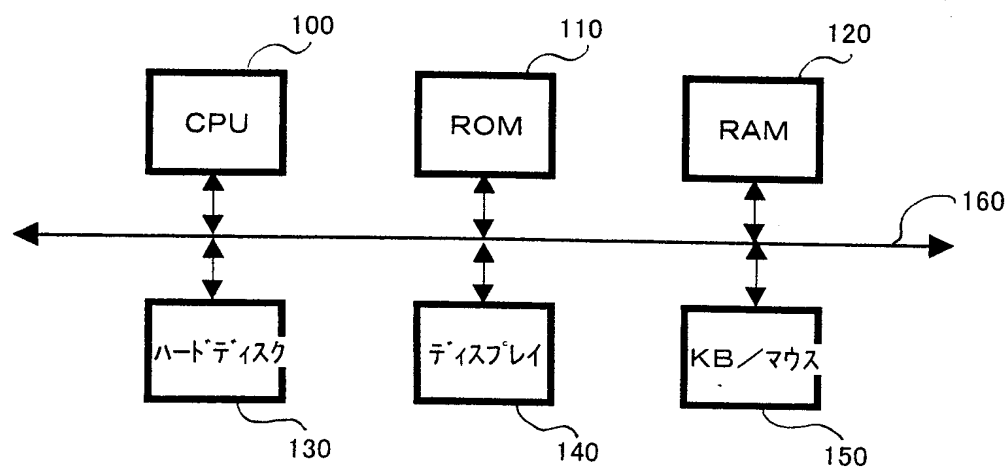


図3

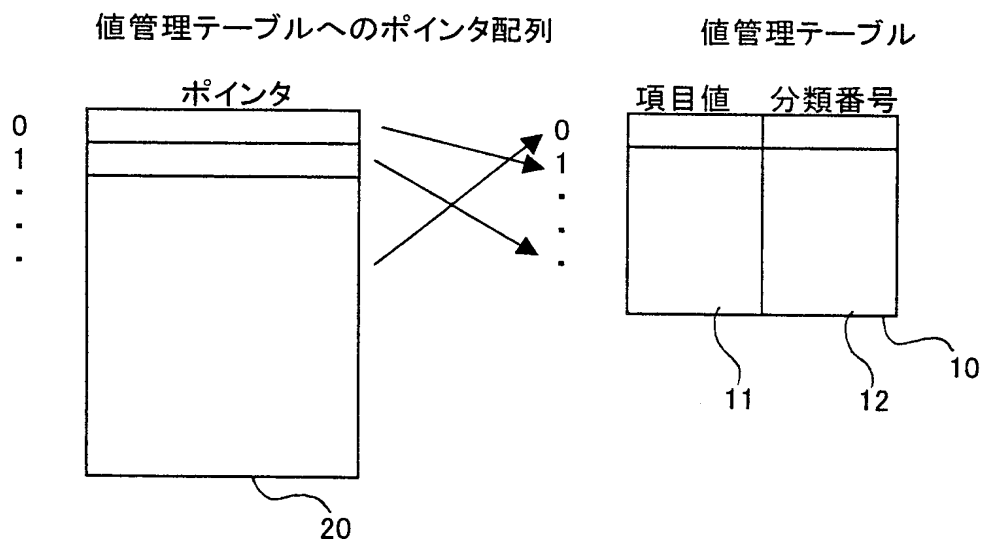


図4

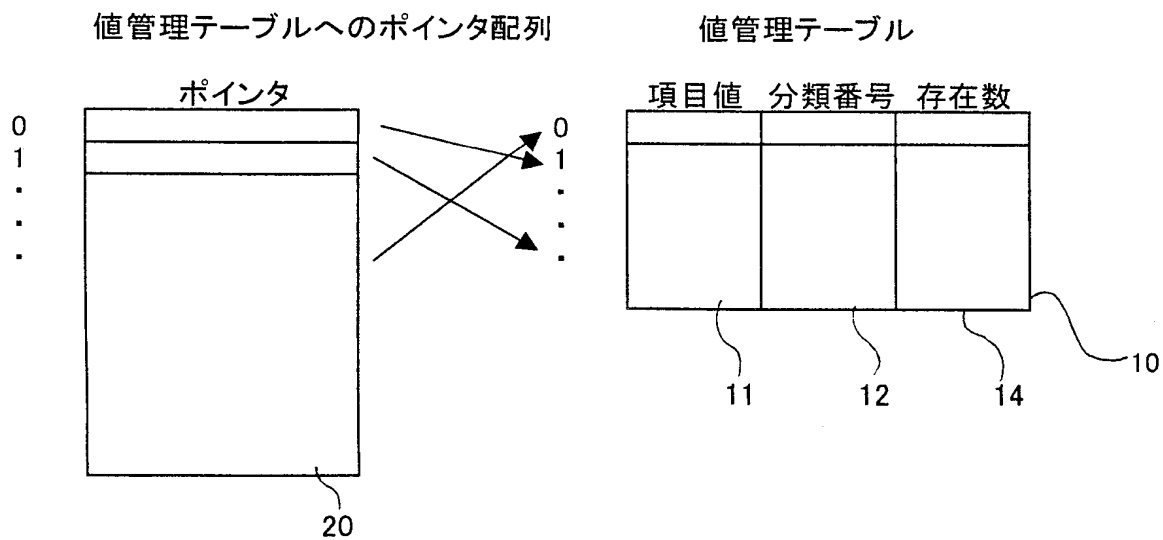


図5

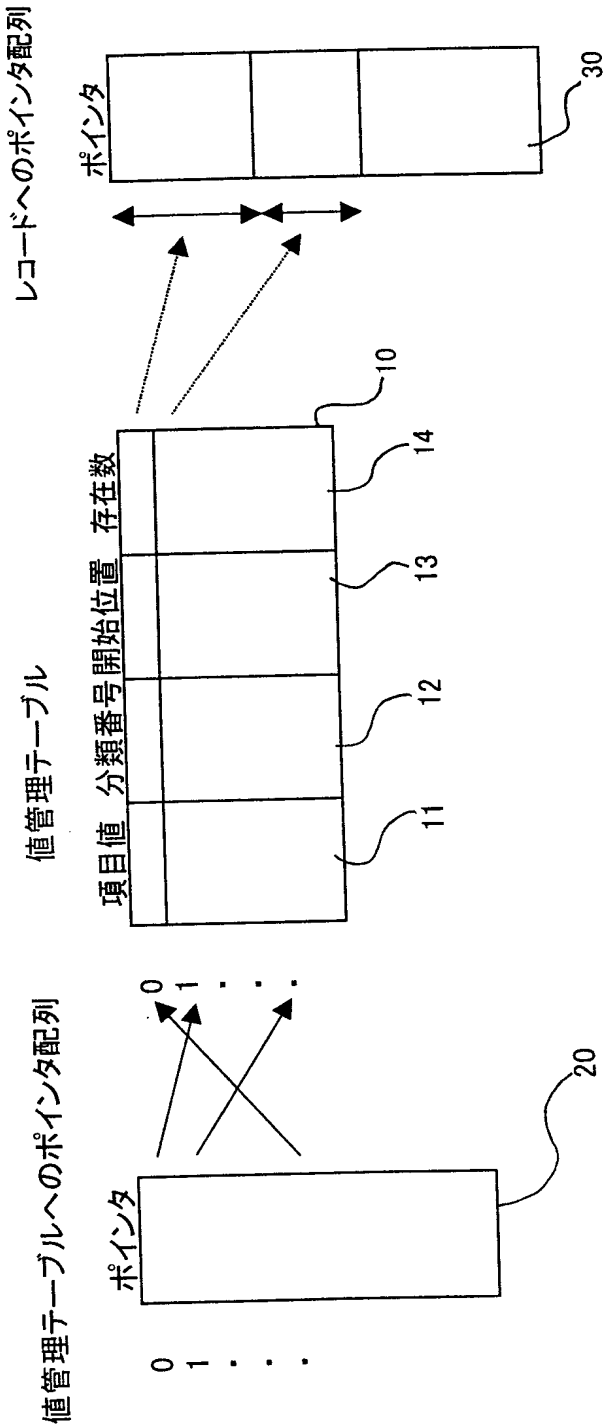


図6

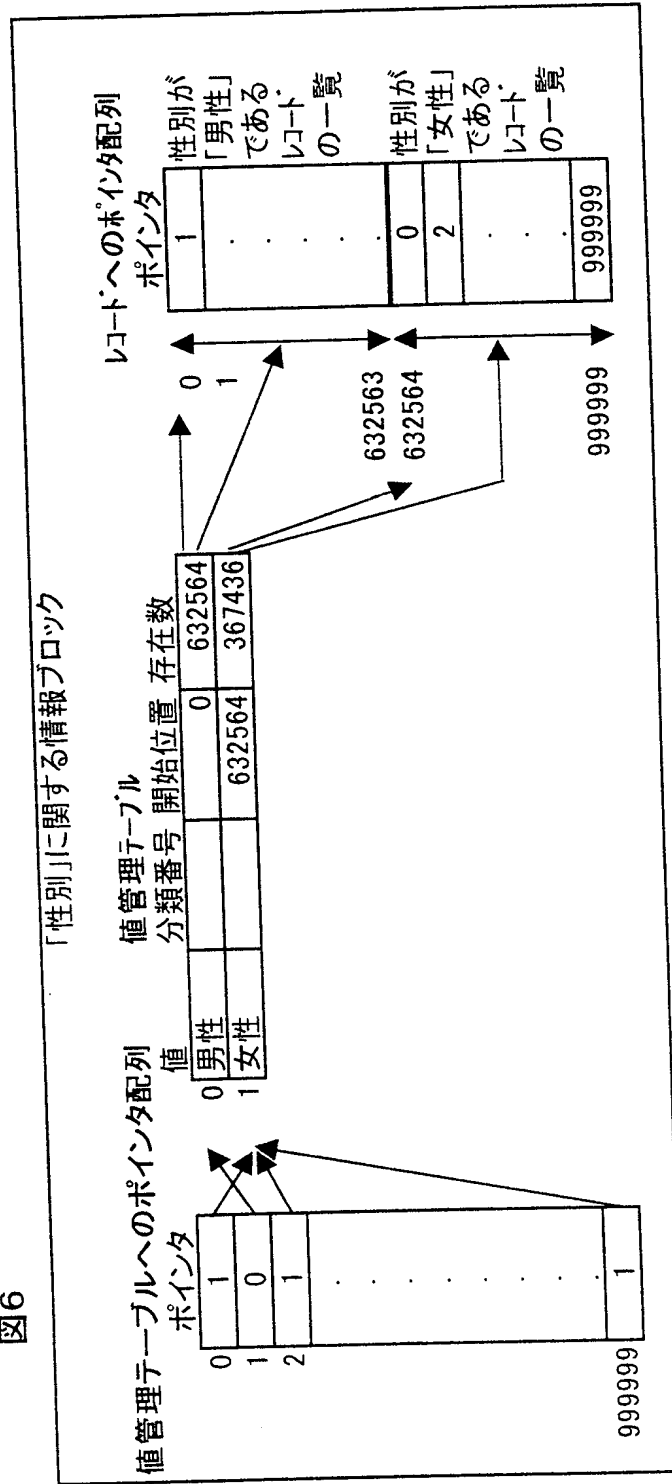
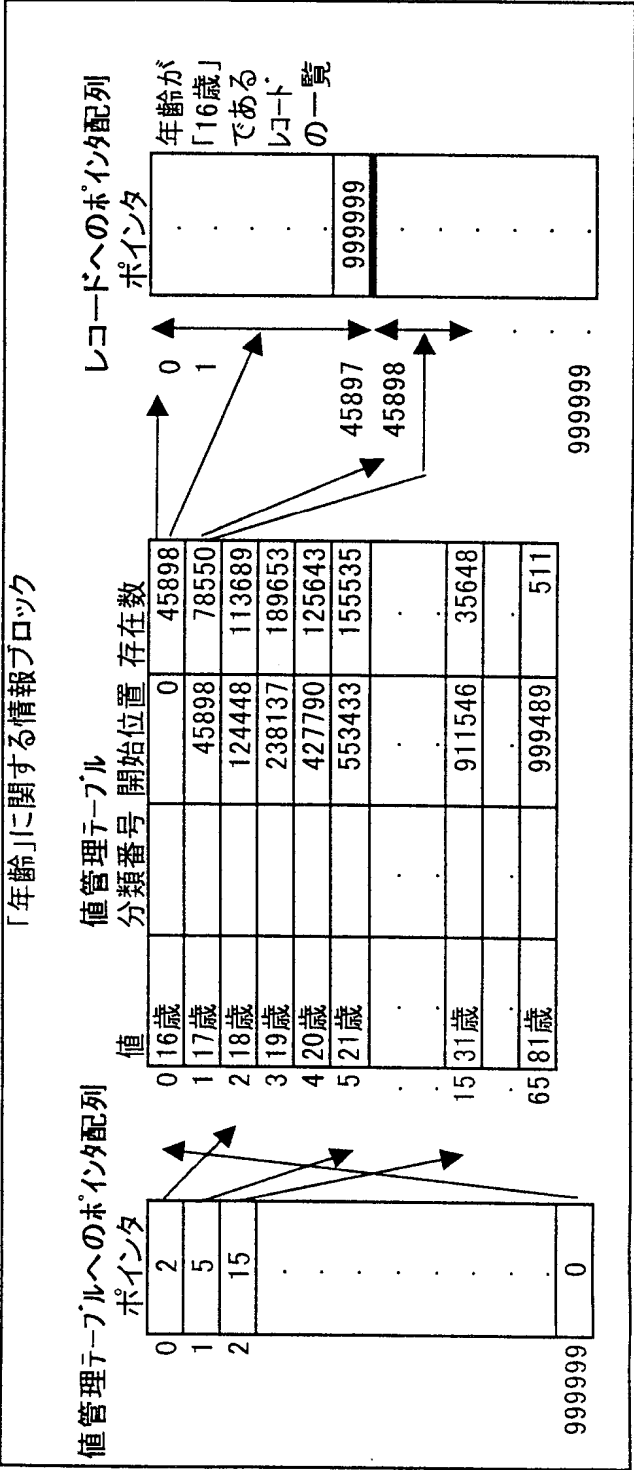


図7



「年齢」に関する情報ブロック

図8

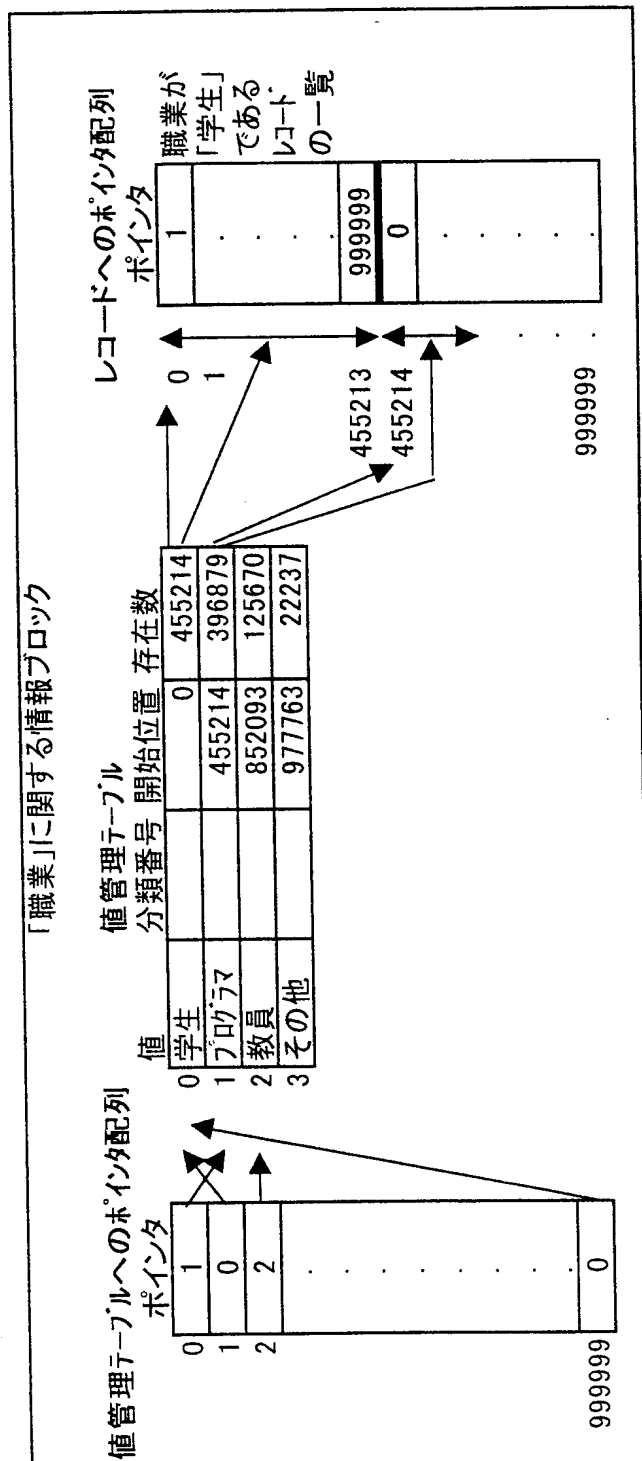


図9

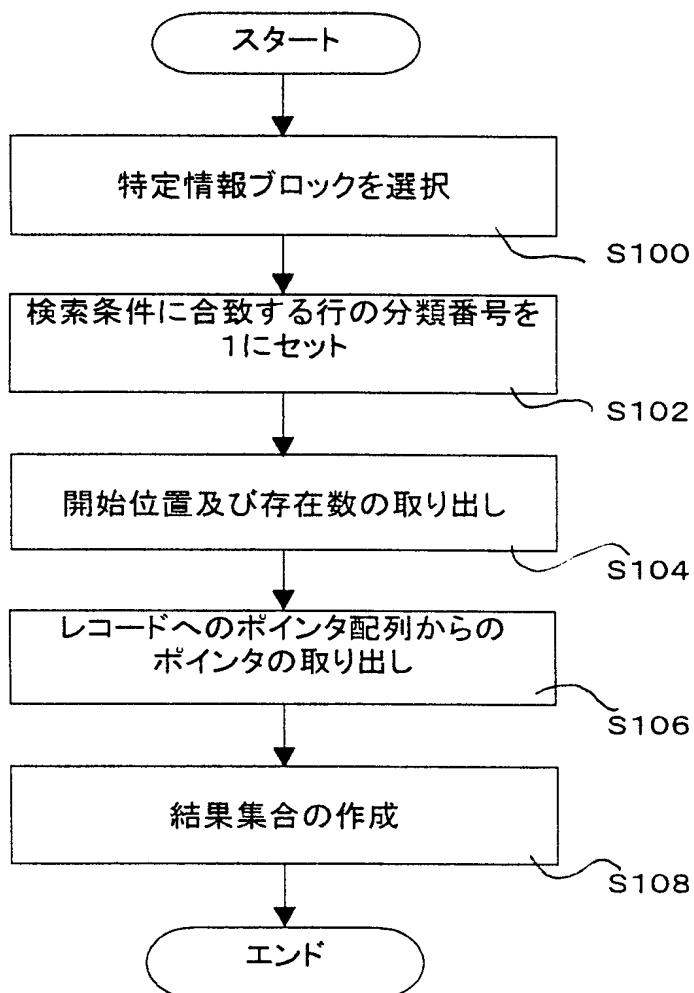


図10

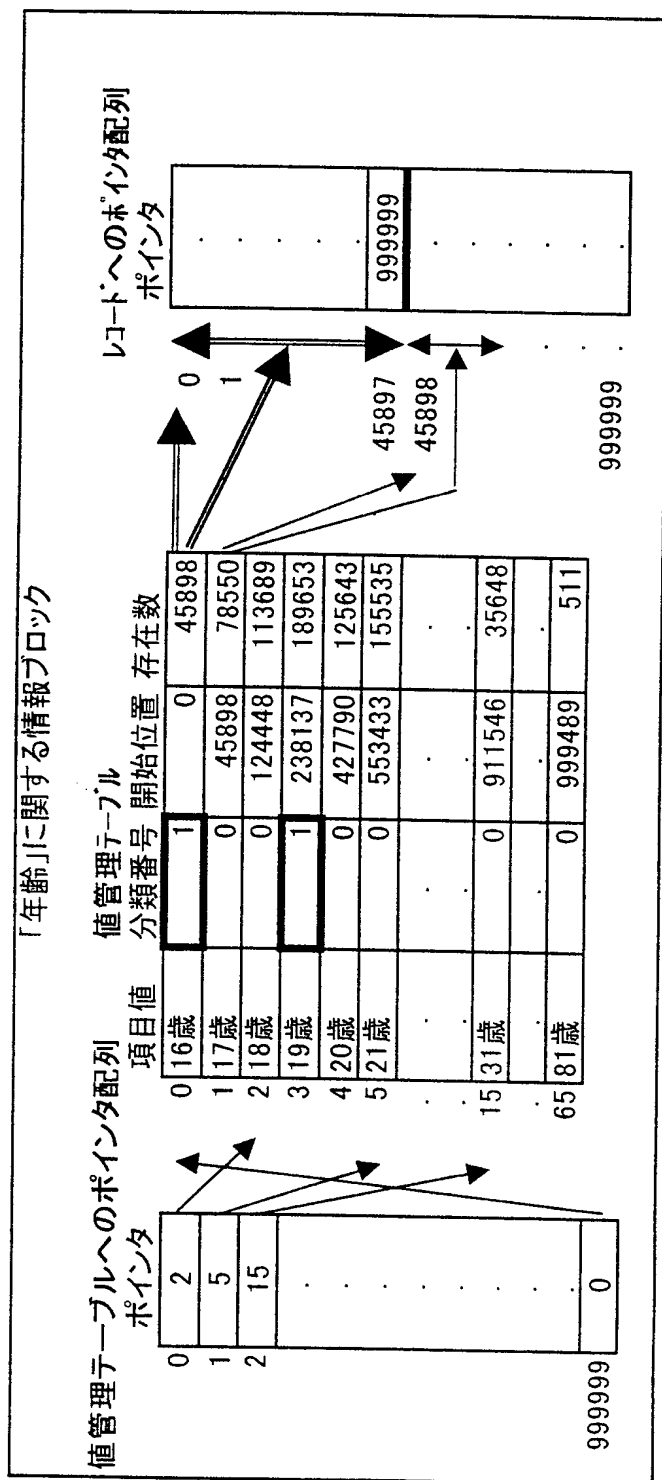




図11

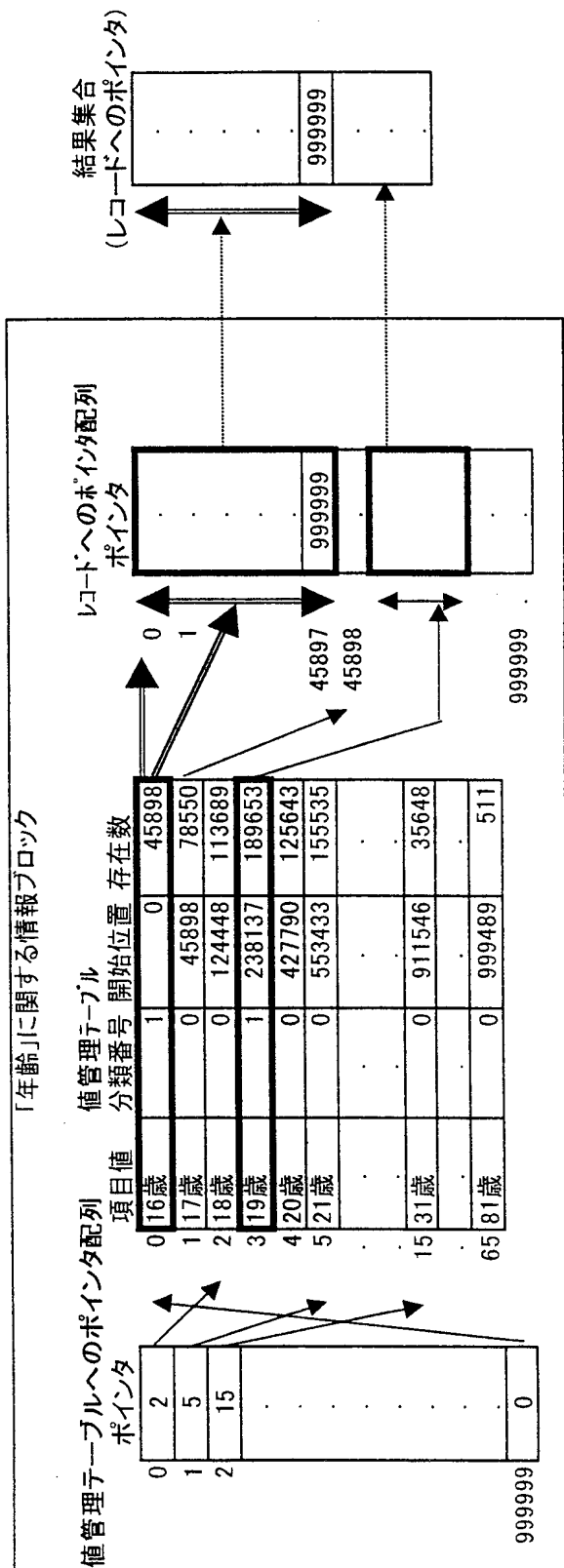
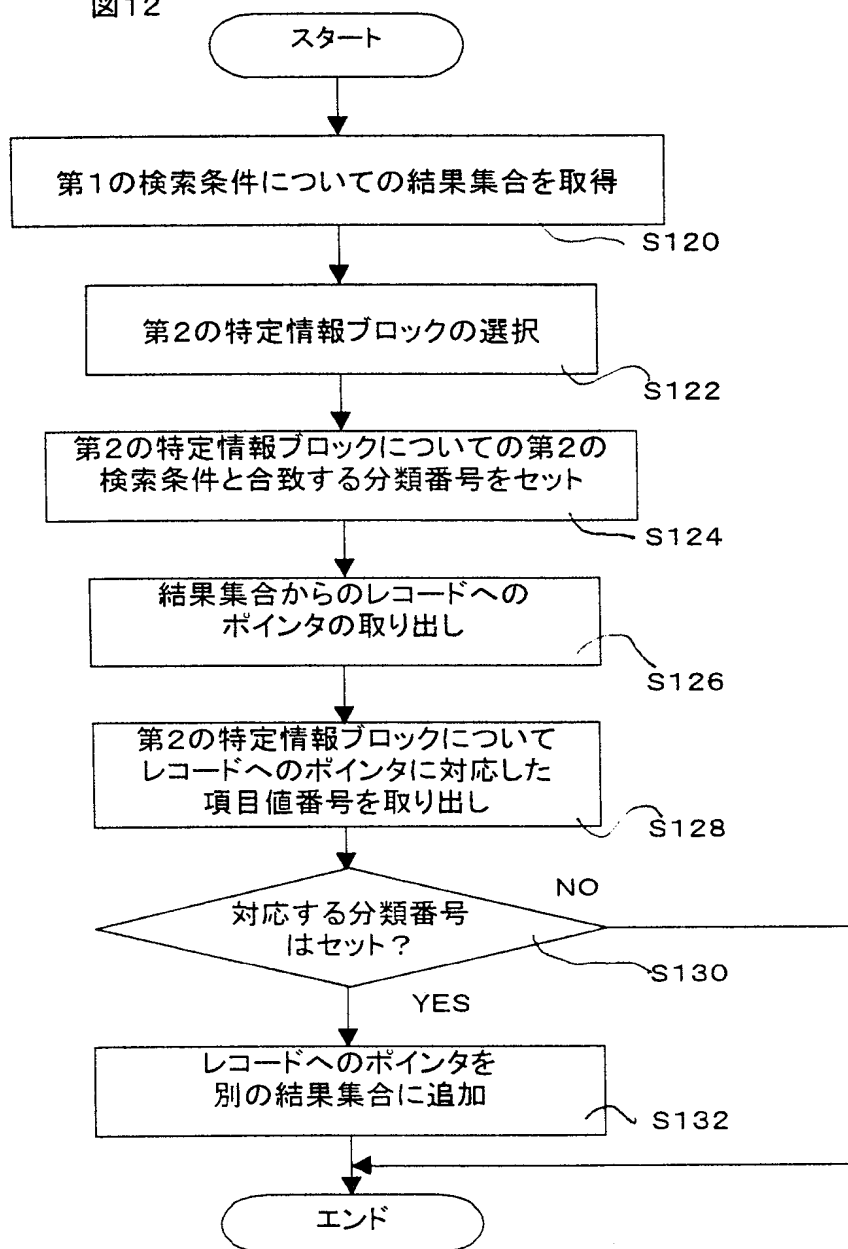


図12



13

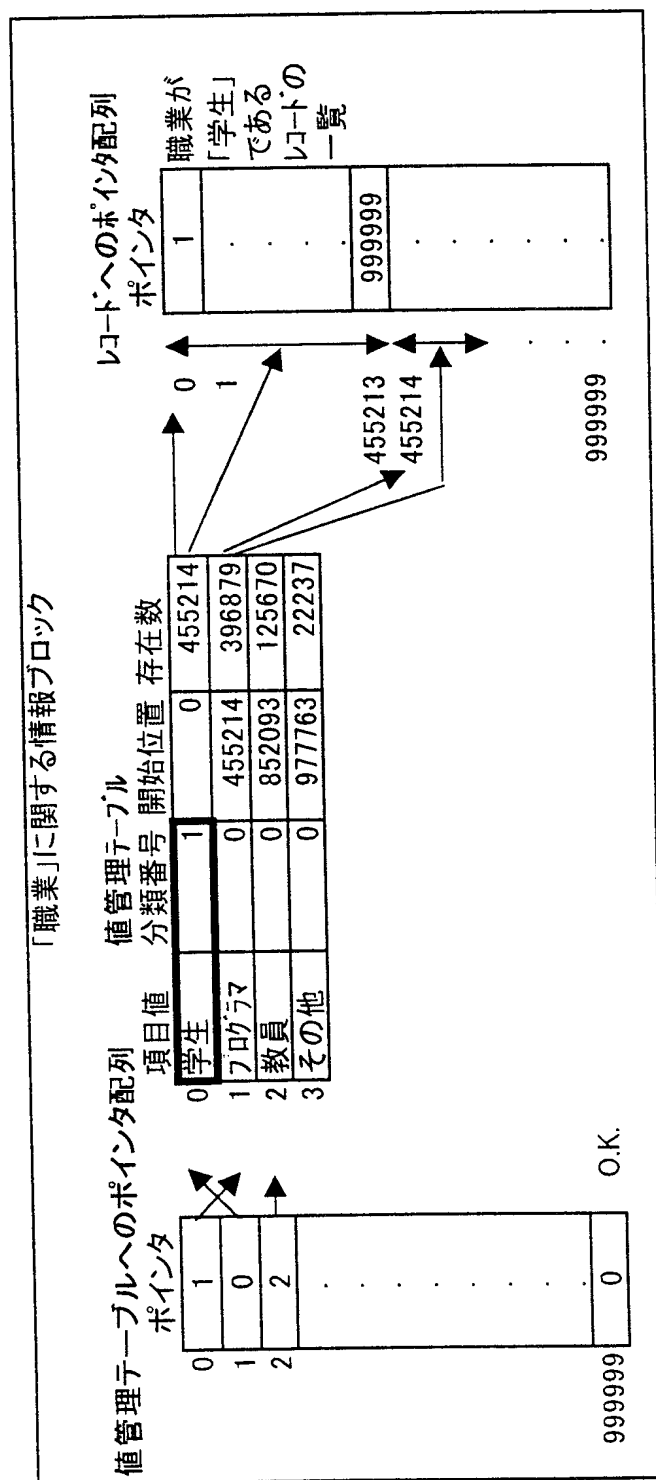




図15

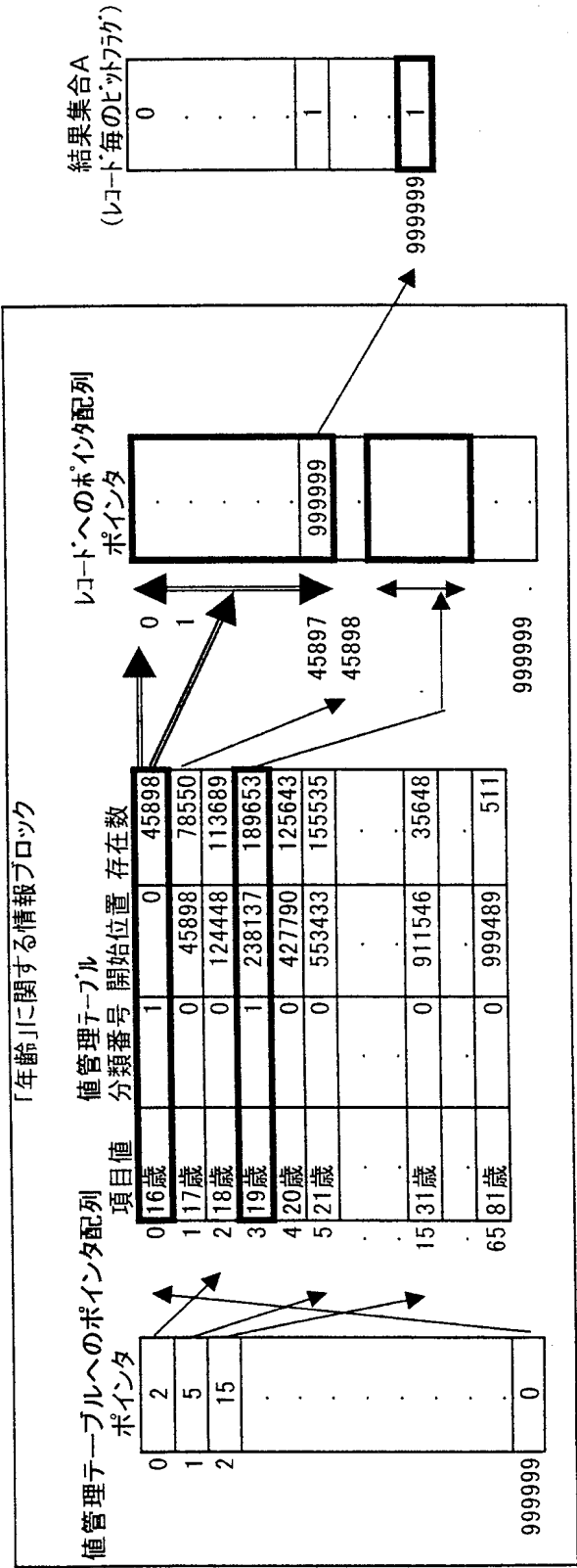


図16

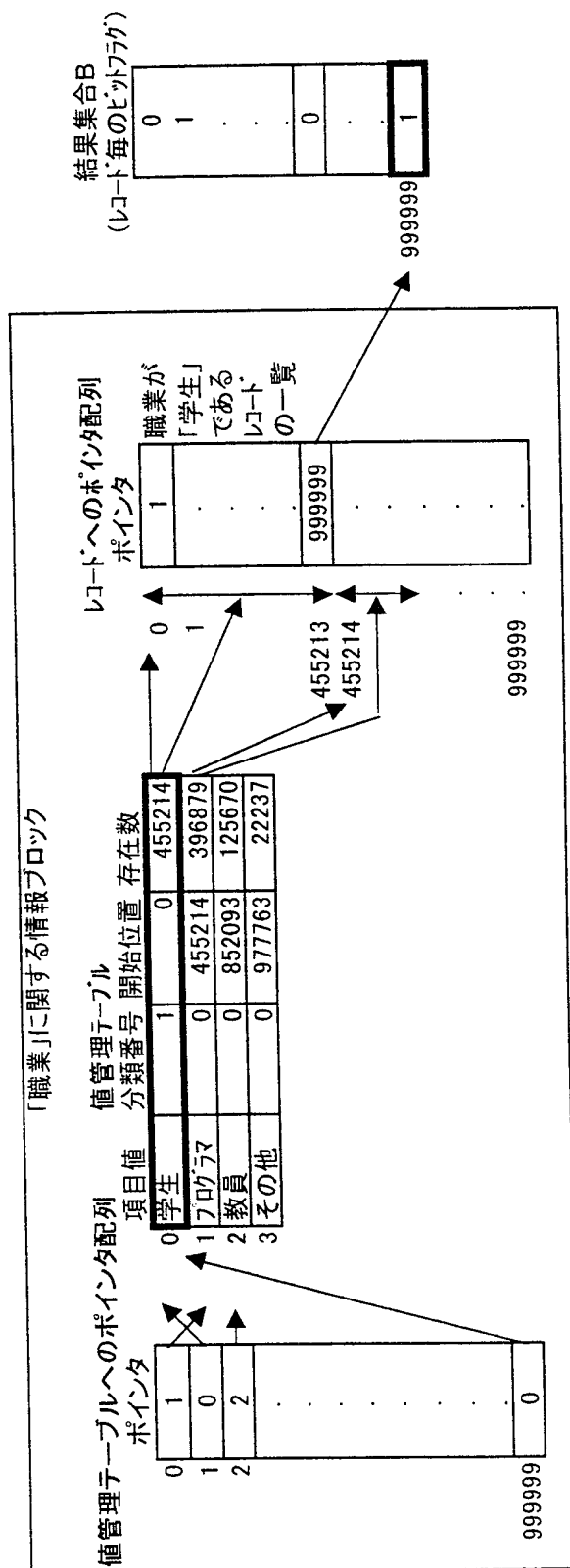


図17

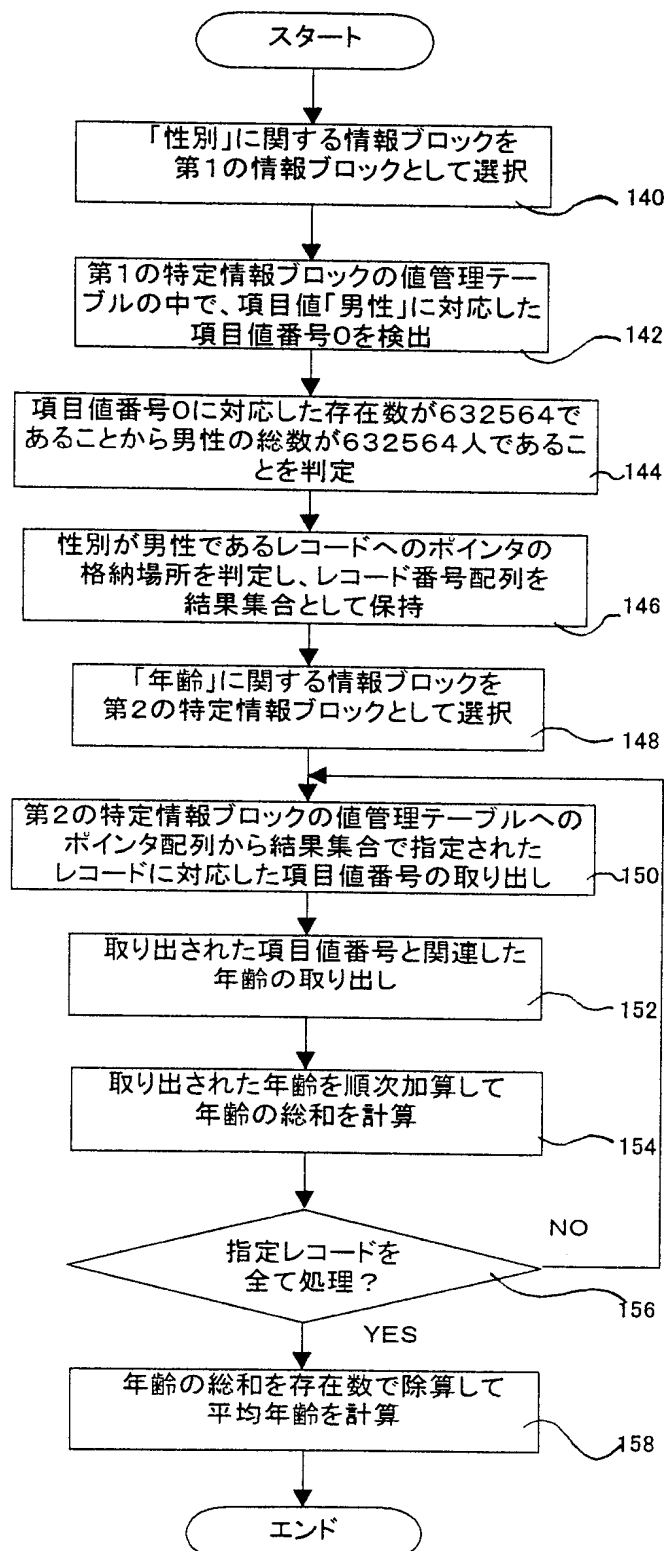


図18

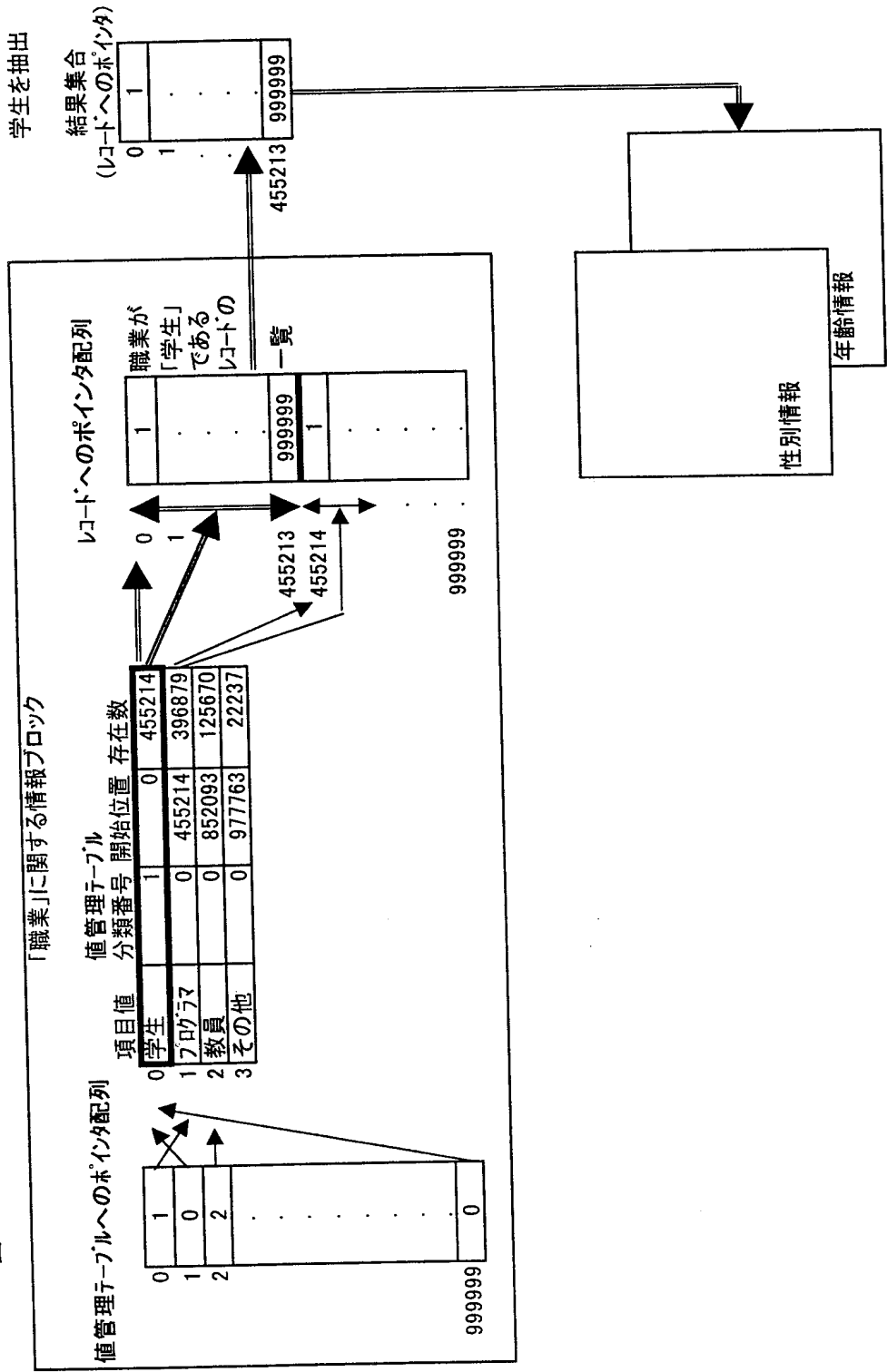




図19

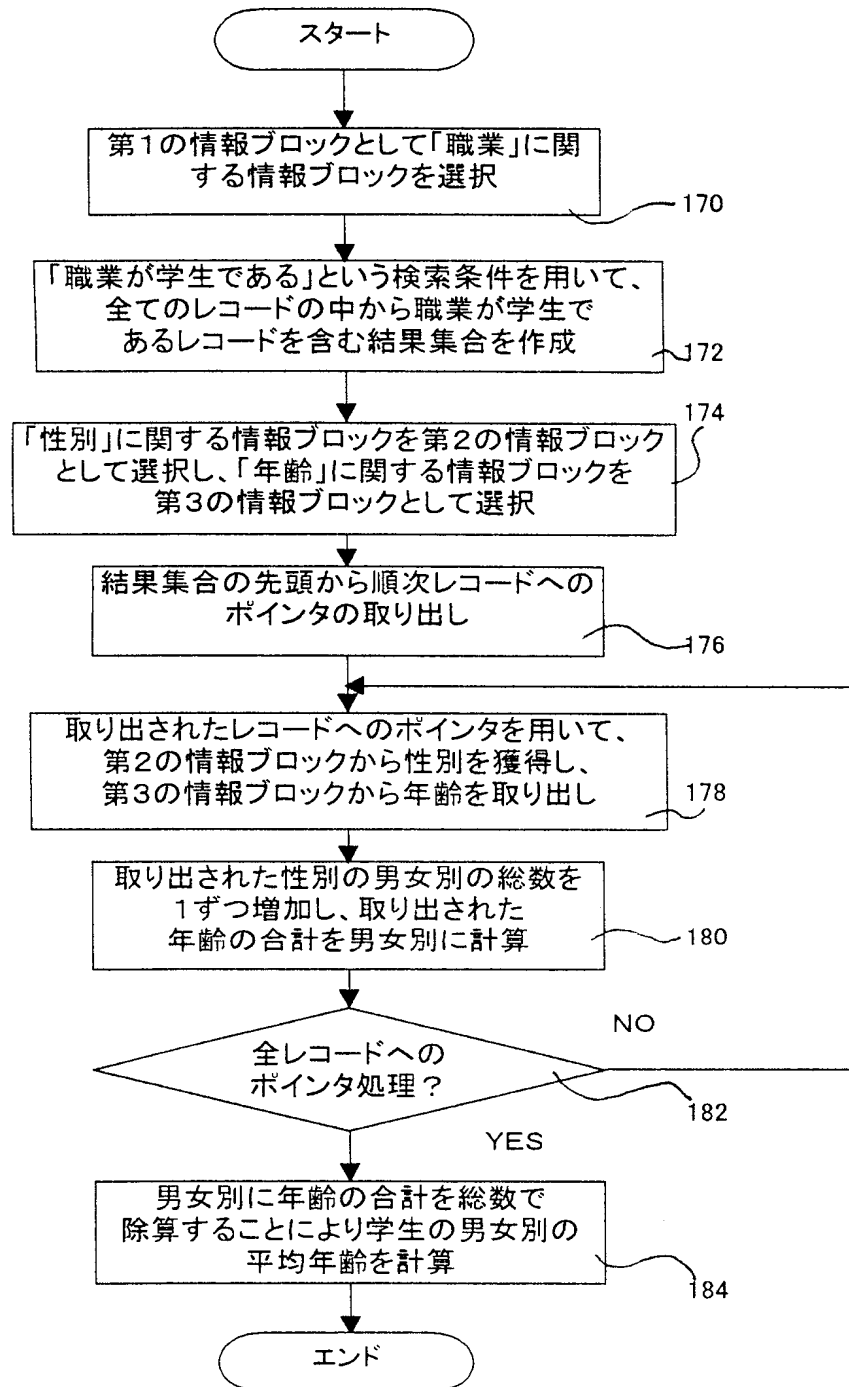


図20

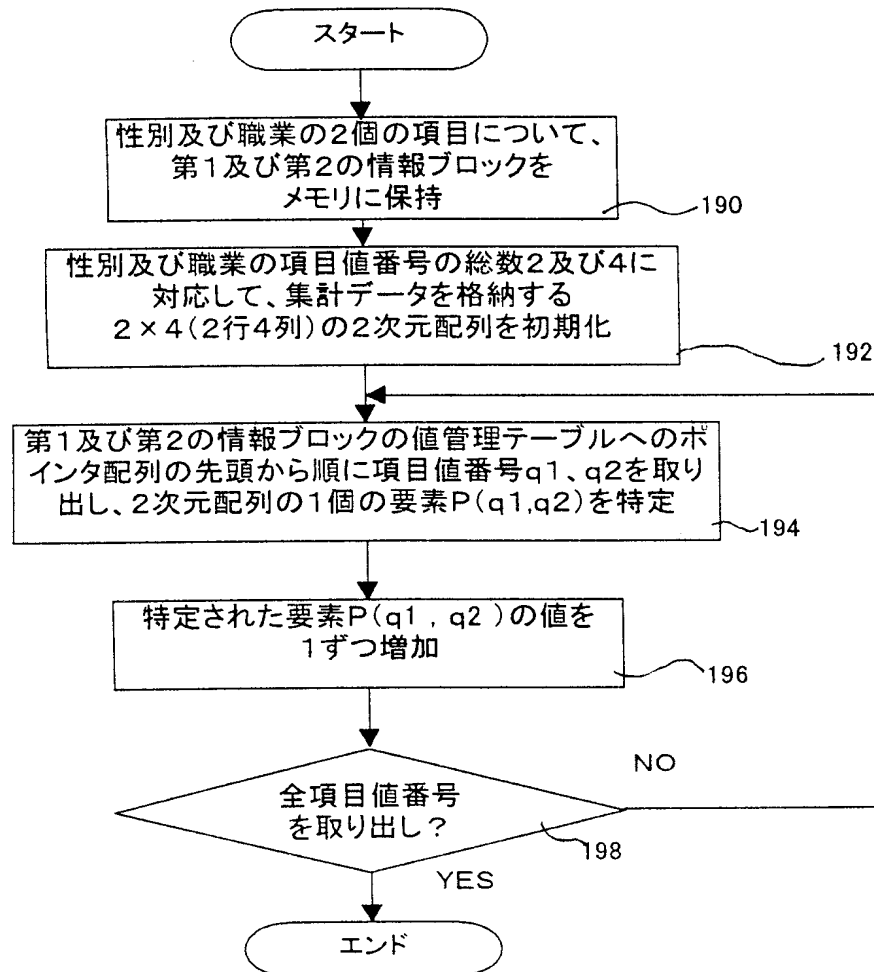


図21

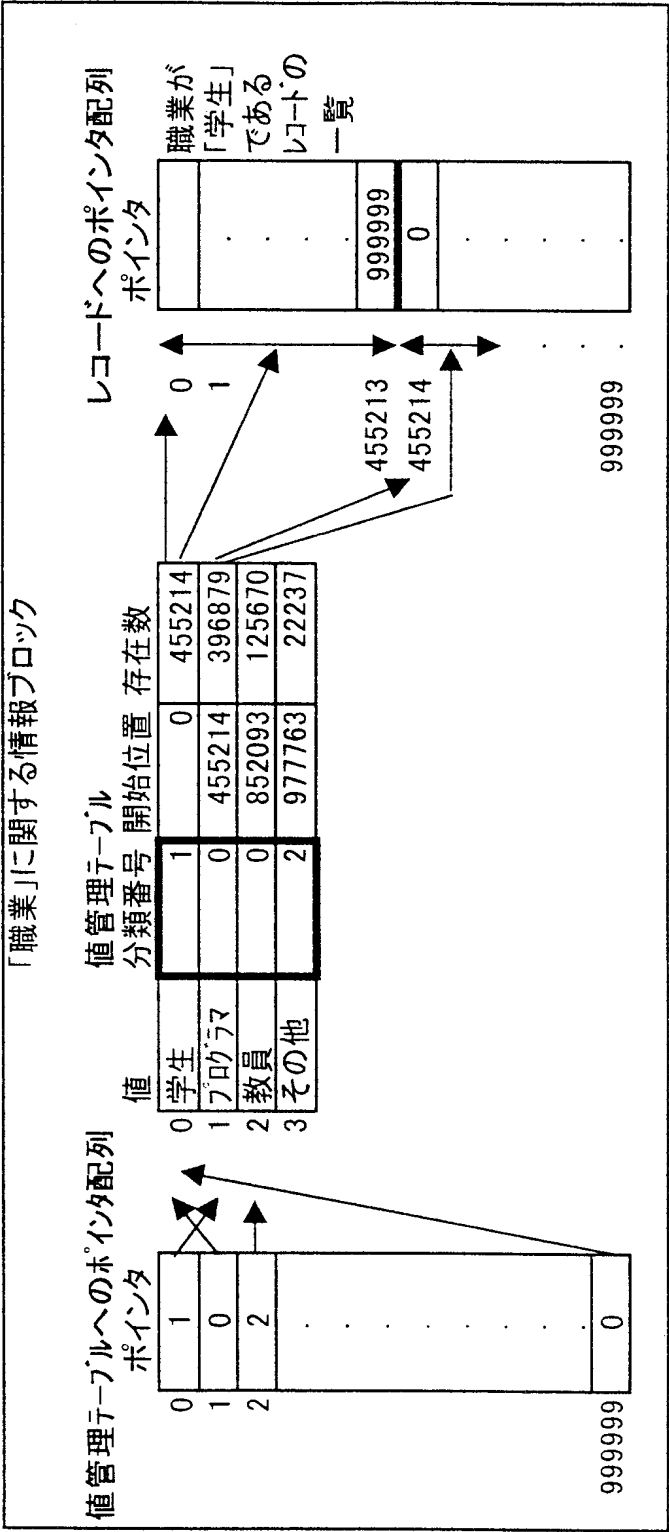


図22

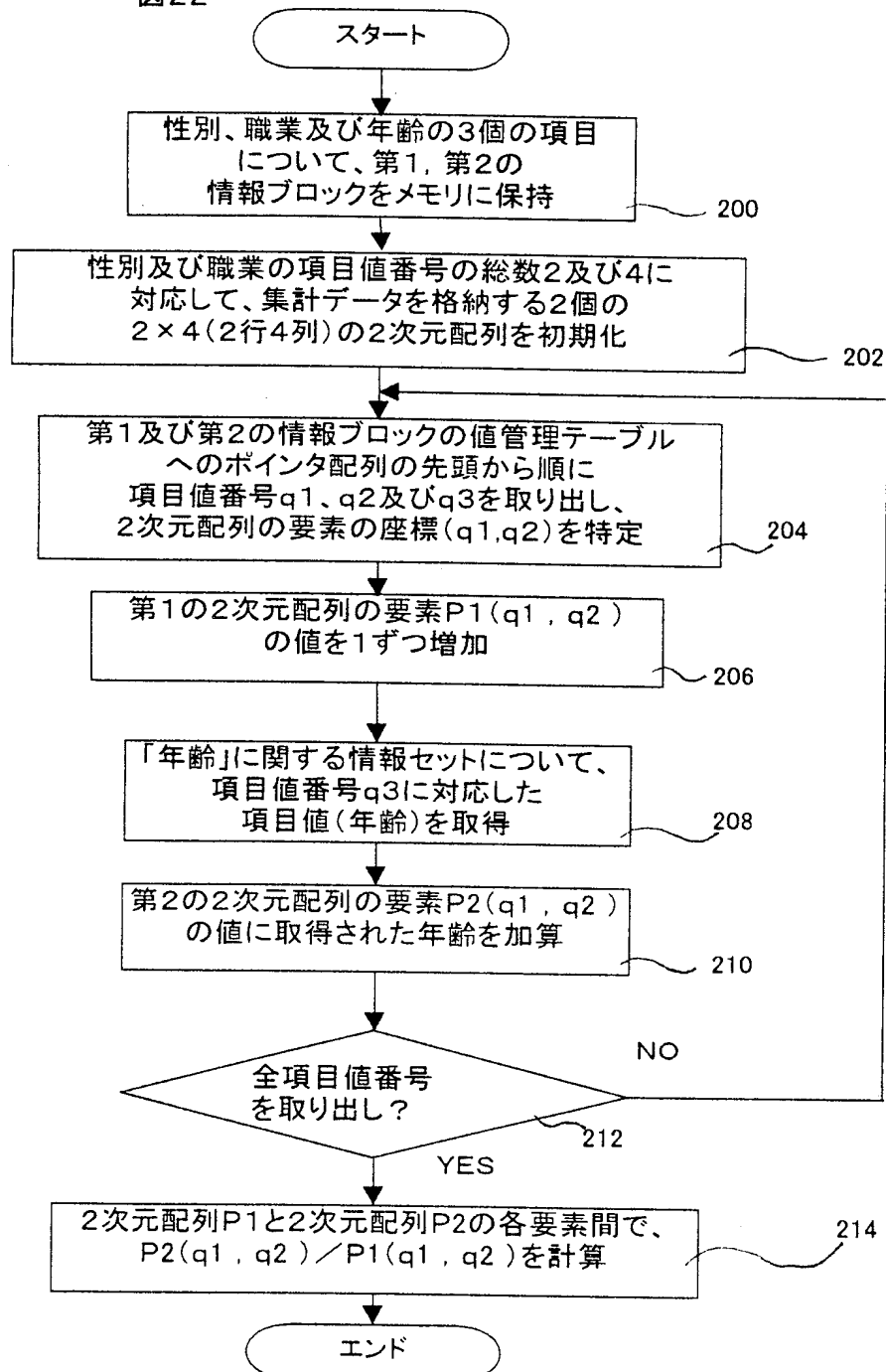


図23

(A)

	学生	プログラマ	教員	無職
男性	○	○	○	○
女性	○	○	○	○

(B)

	学生	プログラマ	教員	無職
男性	×	×	×	×
女性	○	×	×	×

図24

0	鉛筆; 消しゴム
1	定規
2	紙; 消しゴム
999999	紙; 鉛筆

図26

	学生	プログラマ	教員	その他	ブランク
男性	81	5	0	5	6
女性	54	8	56	6	8
ブランク	3	2	2	5	12

図25

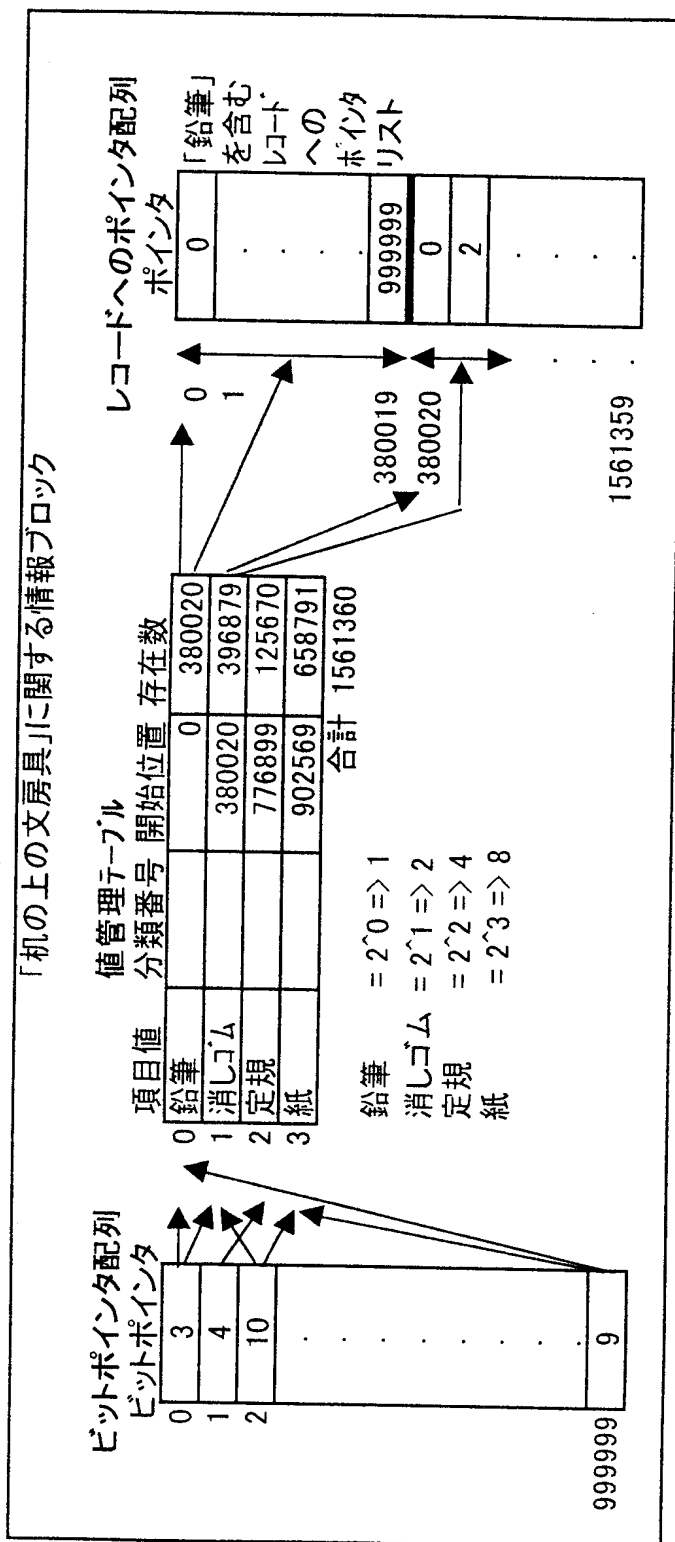


図27

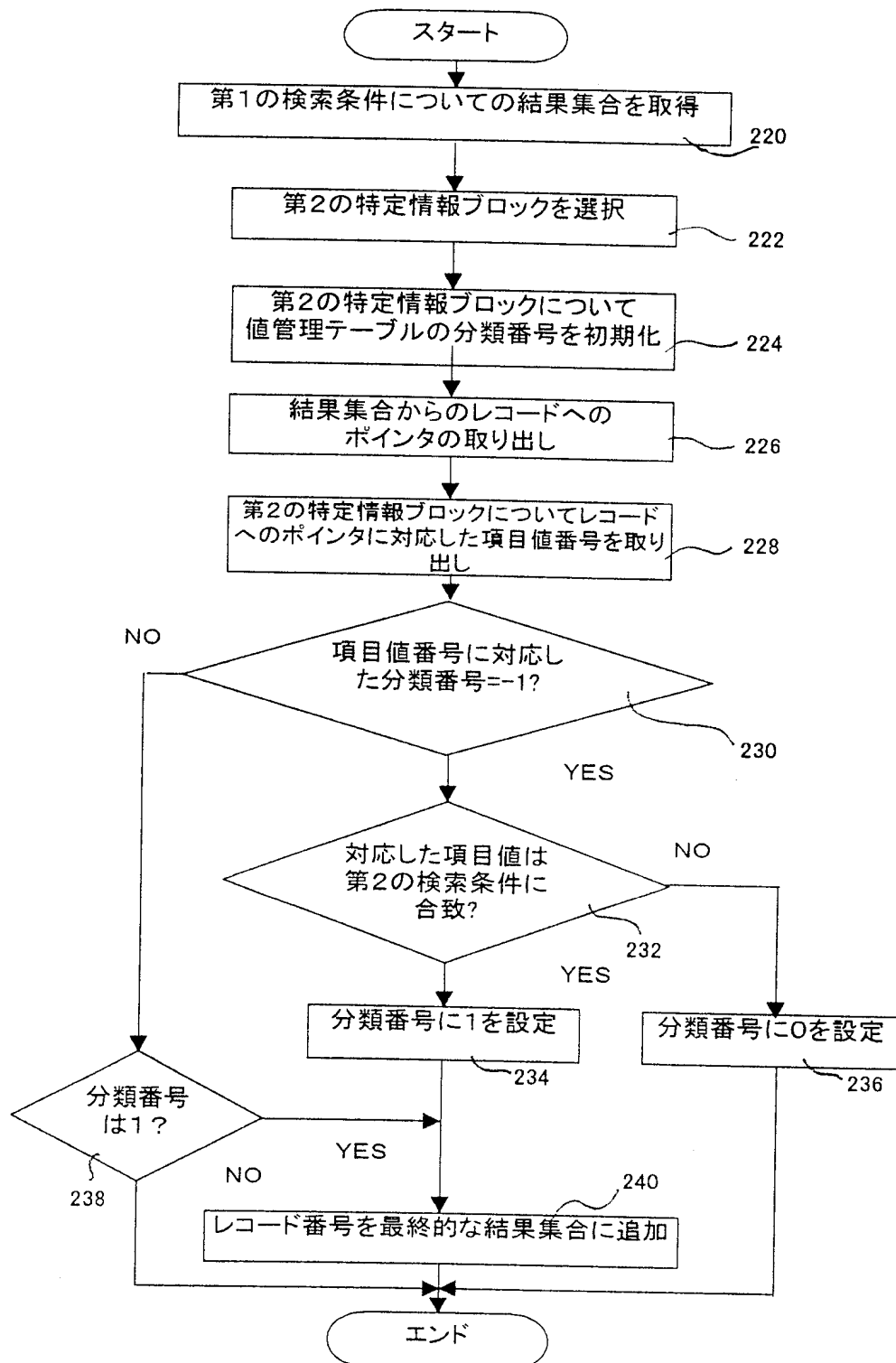
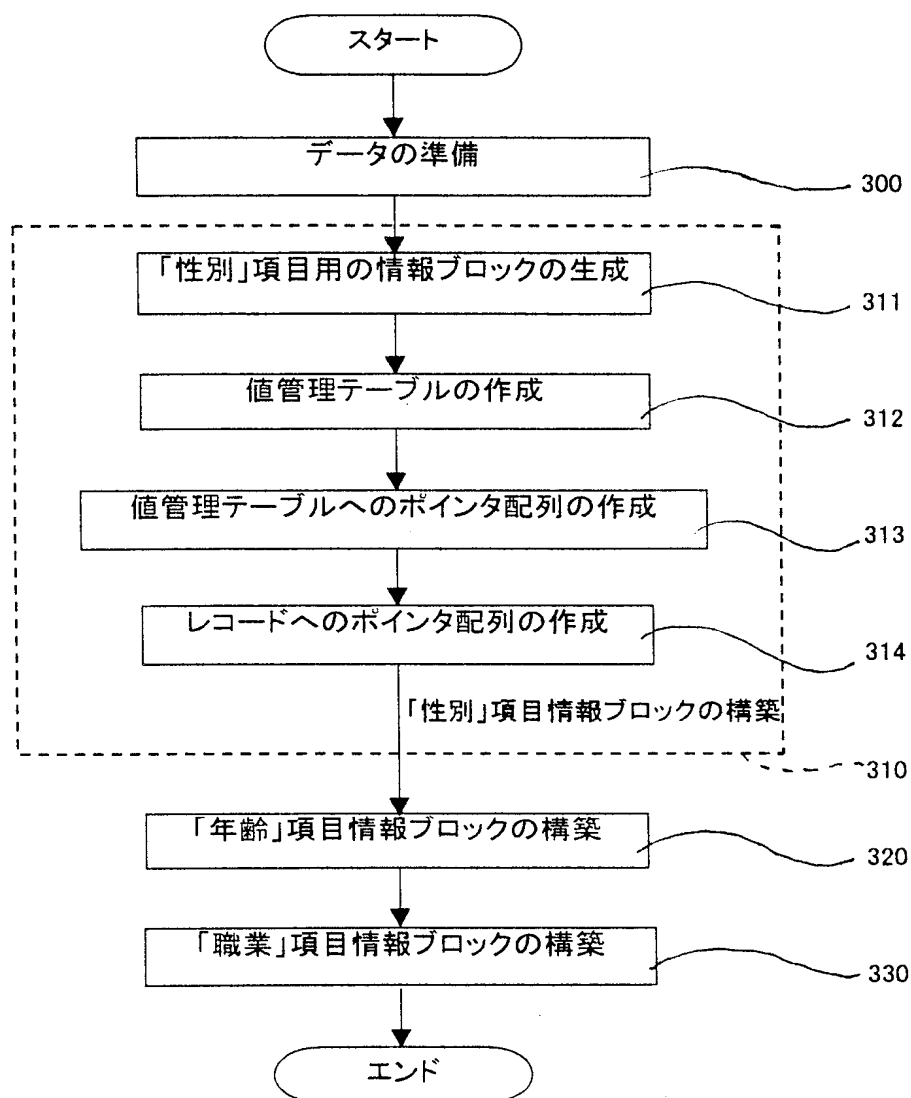
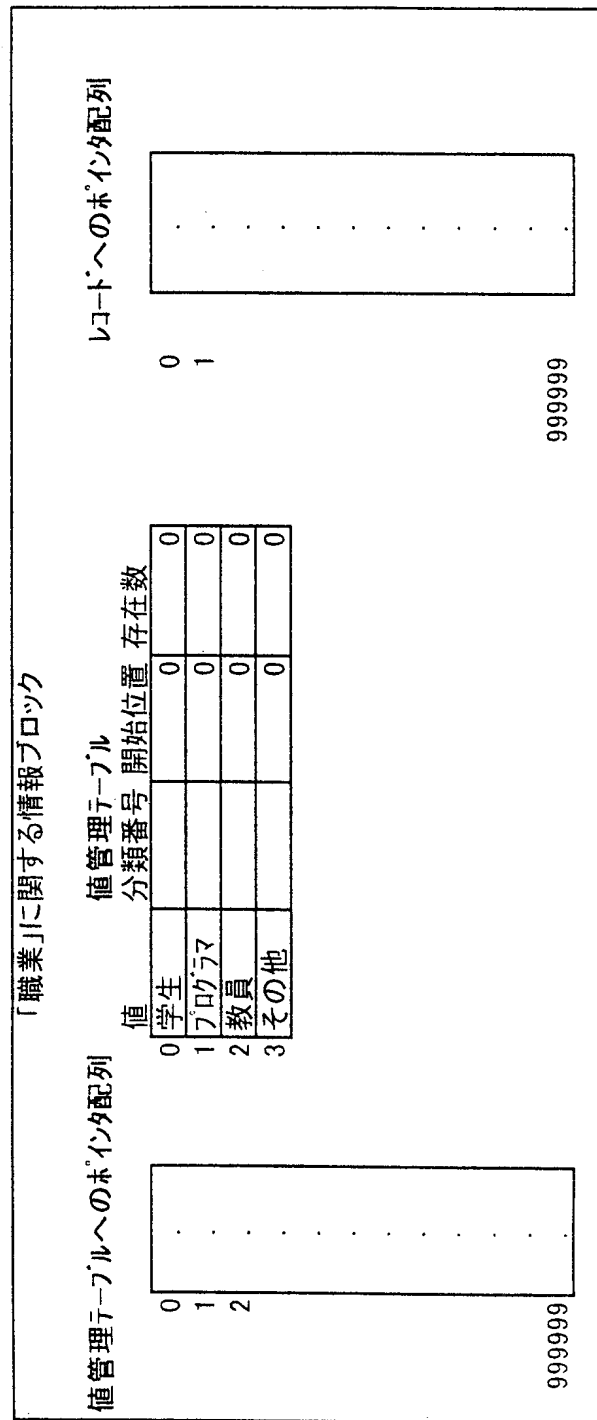


図29





30



投入データ

フ・ロク・ヲマ

学生

員教

•

•

•

•

•

•

1

學生

図 31

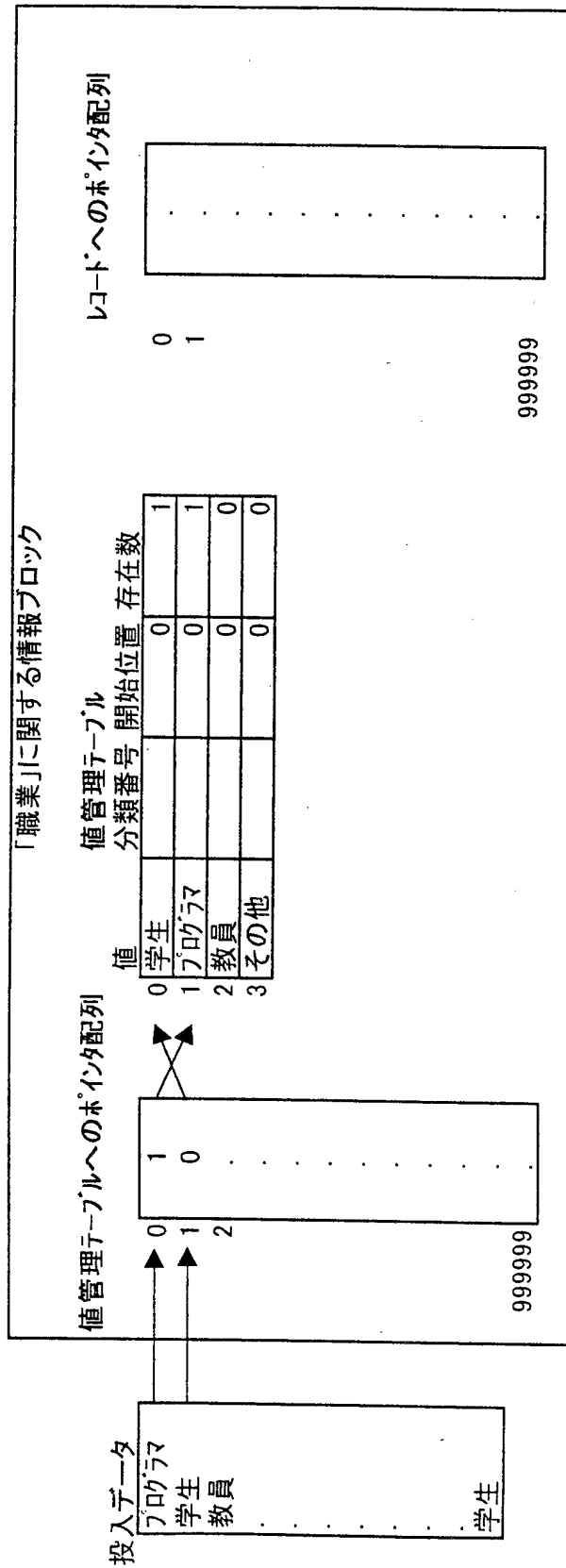
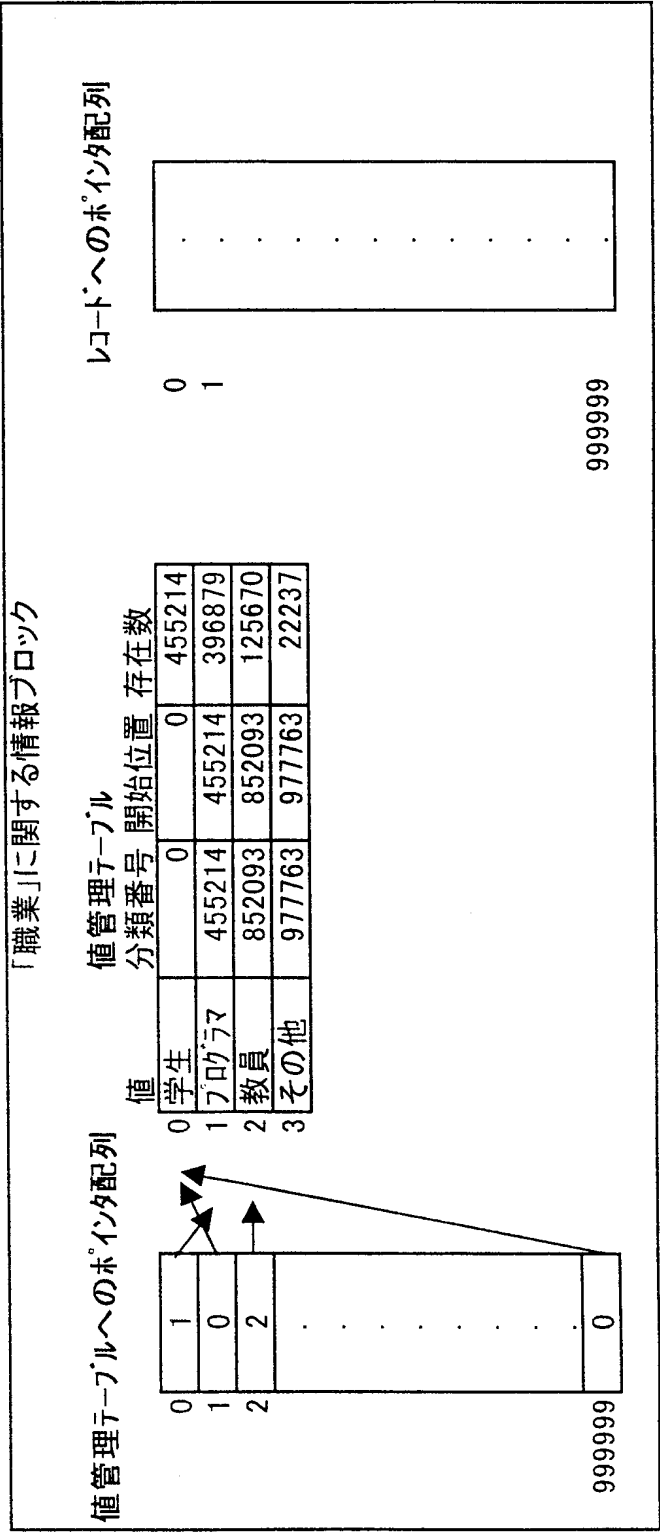
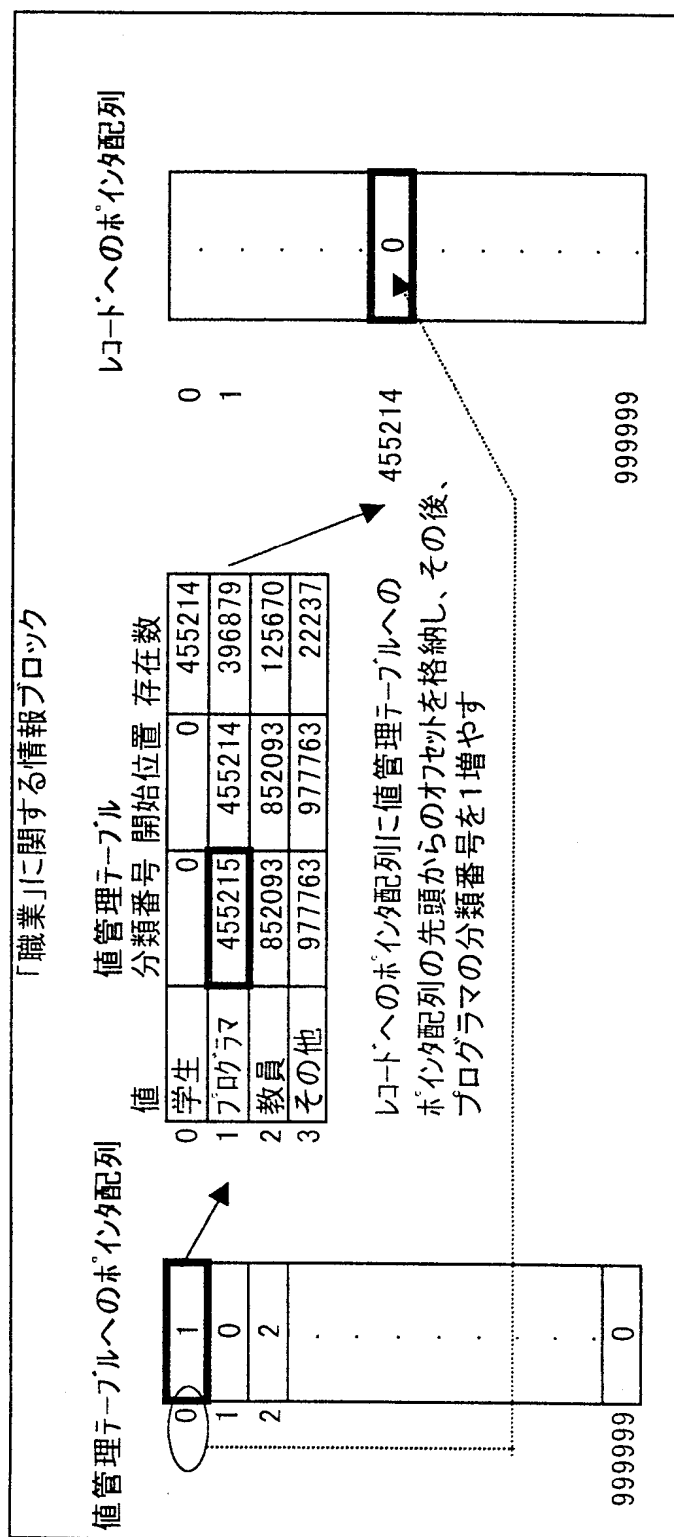


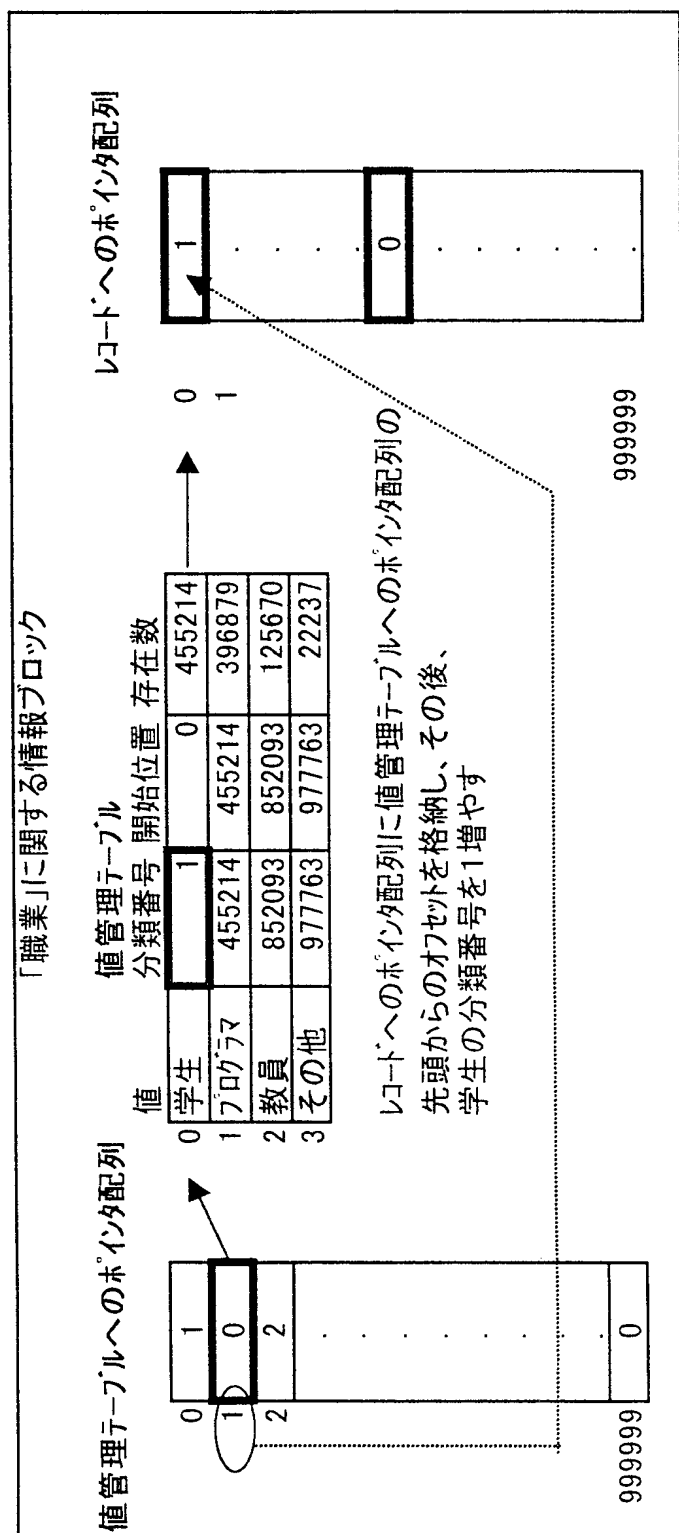
図32



33X



34



53

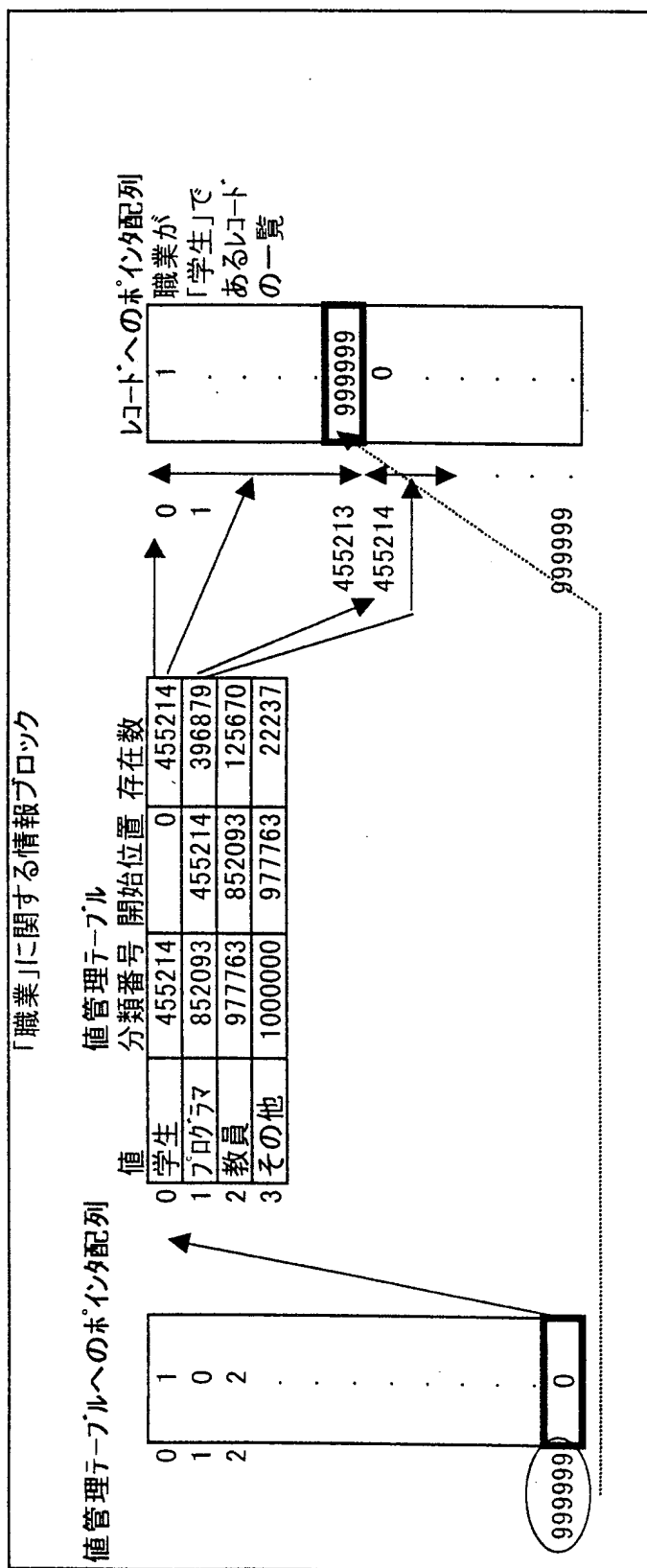


図36

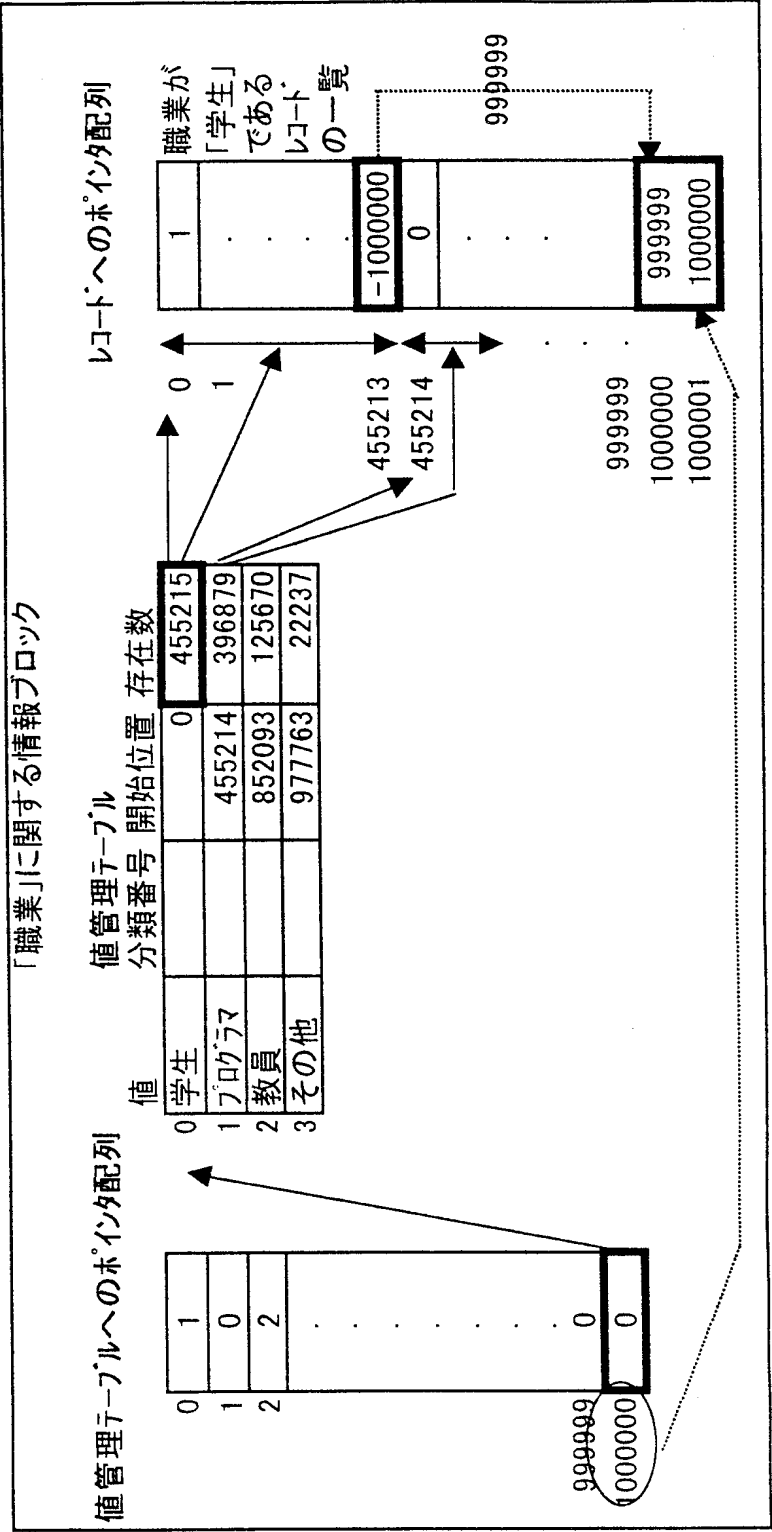


図37

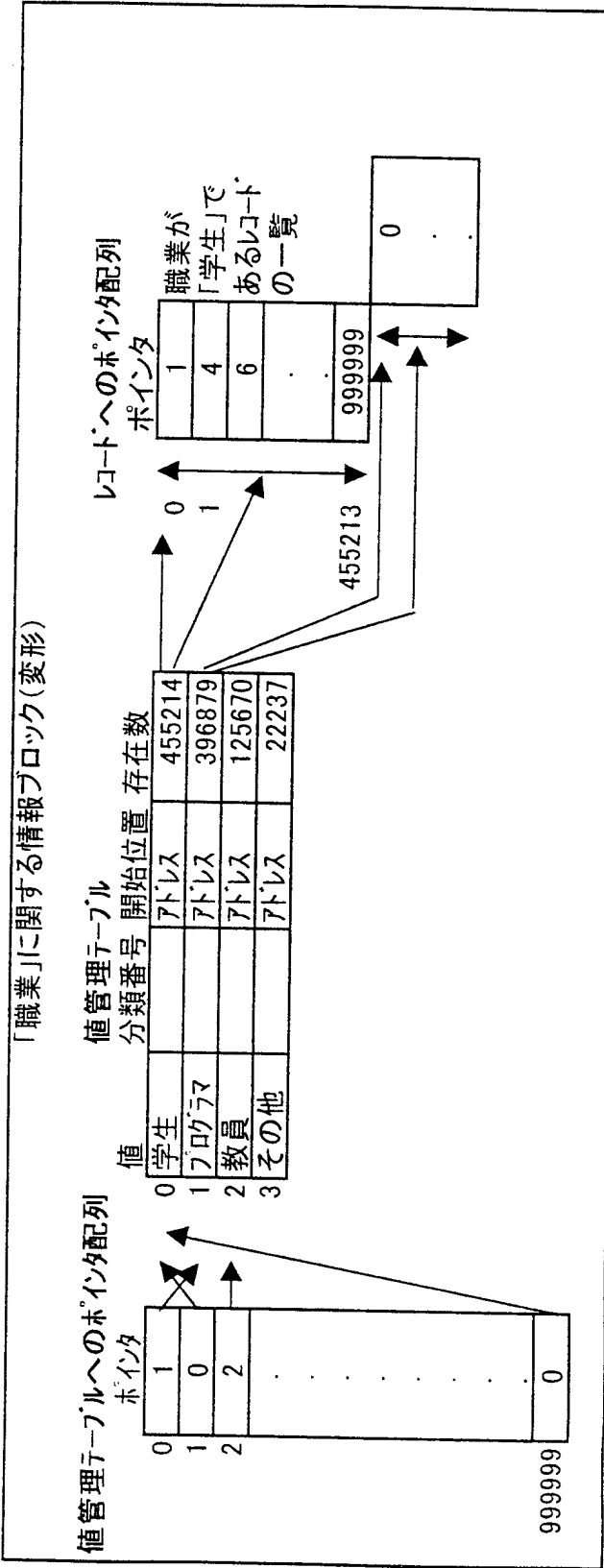




図38

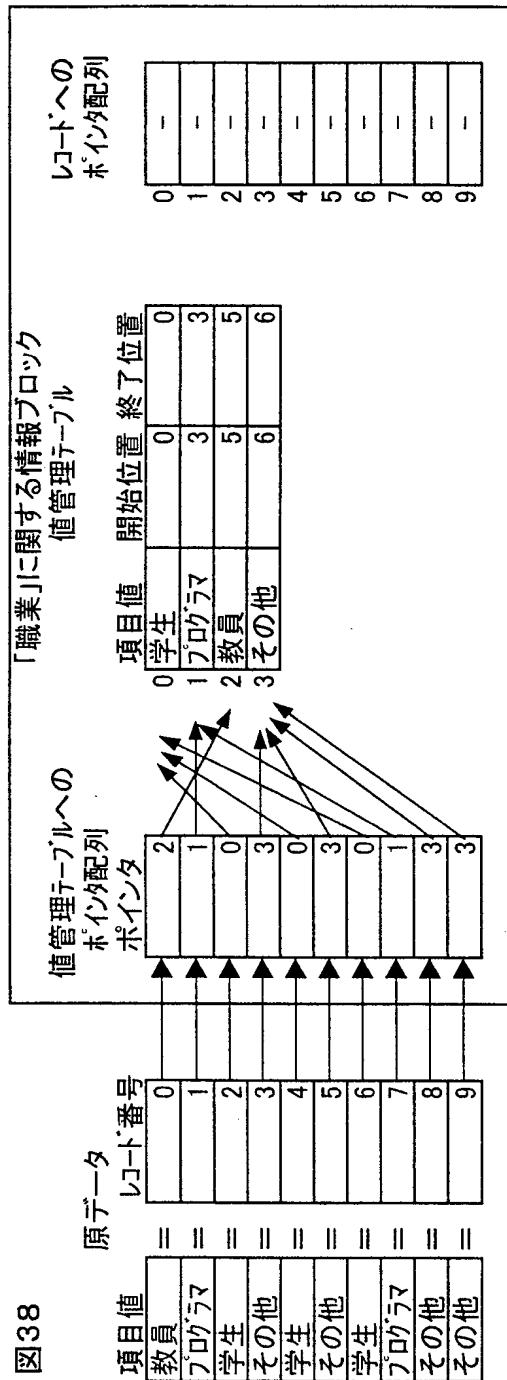


図39

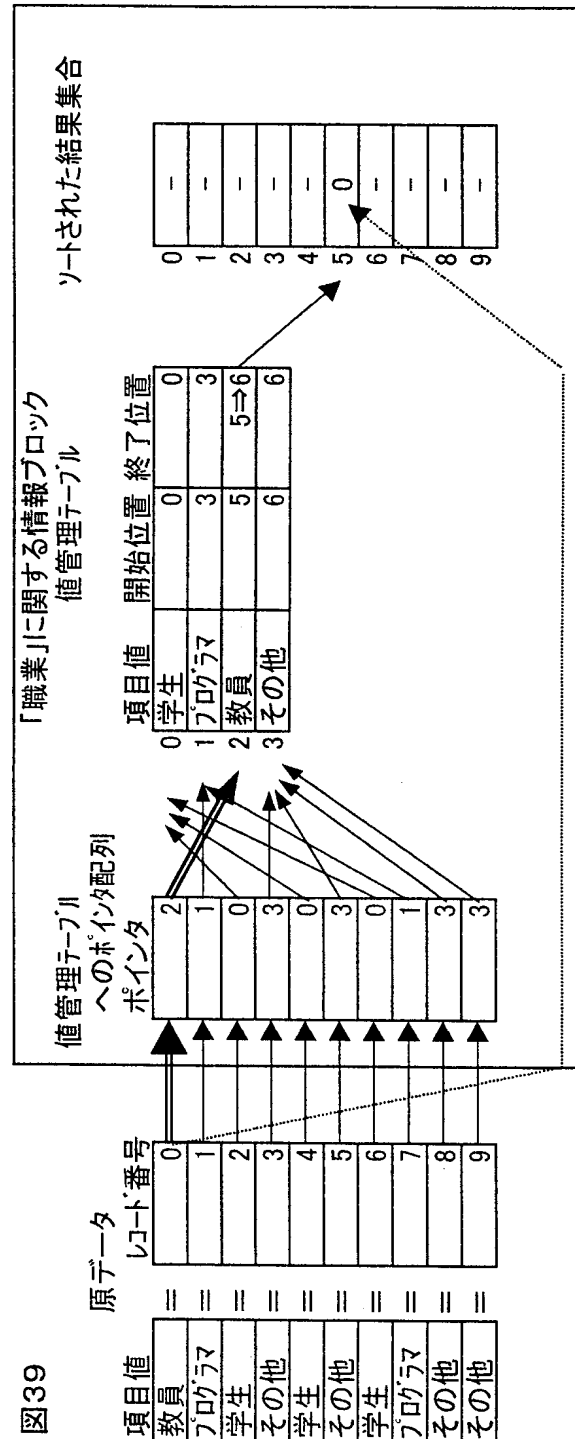


図40

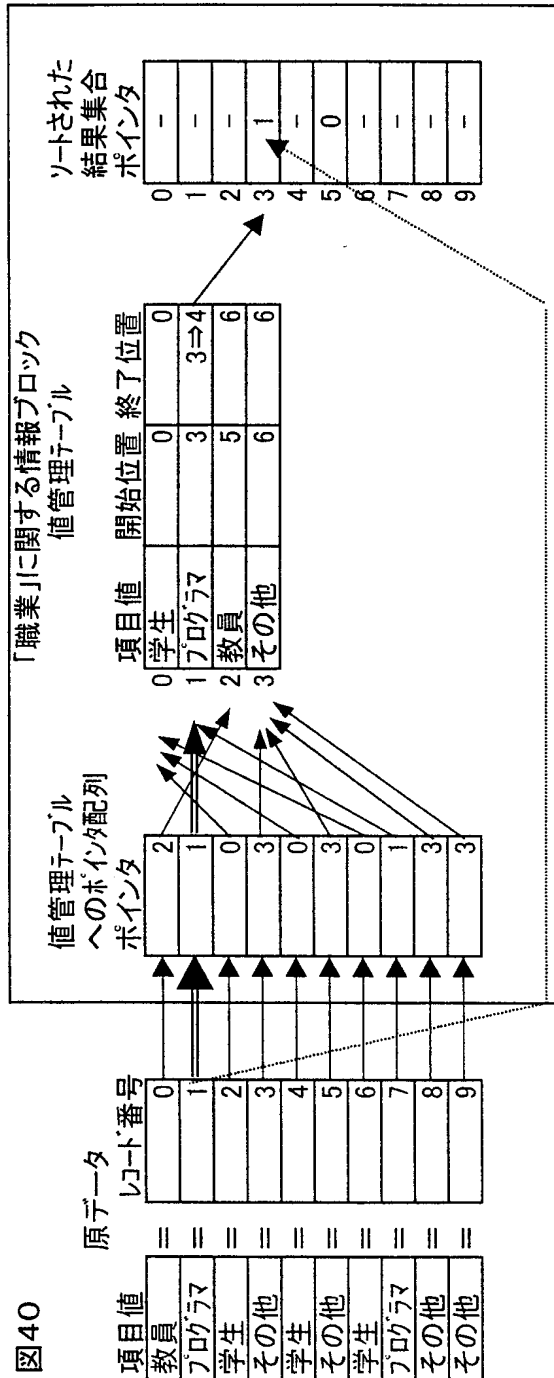
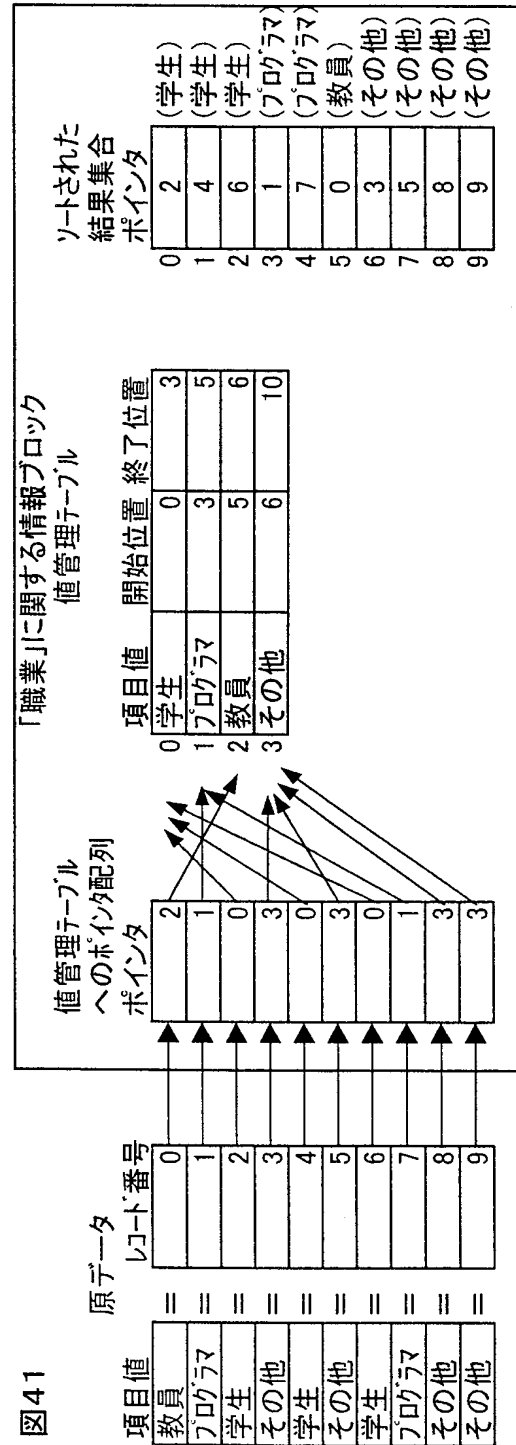


図41



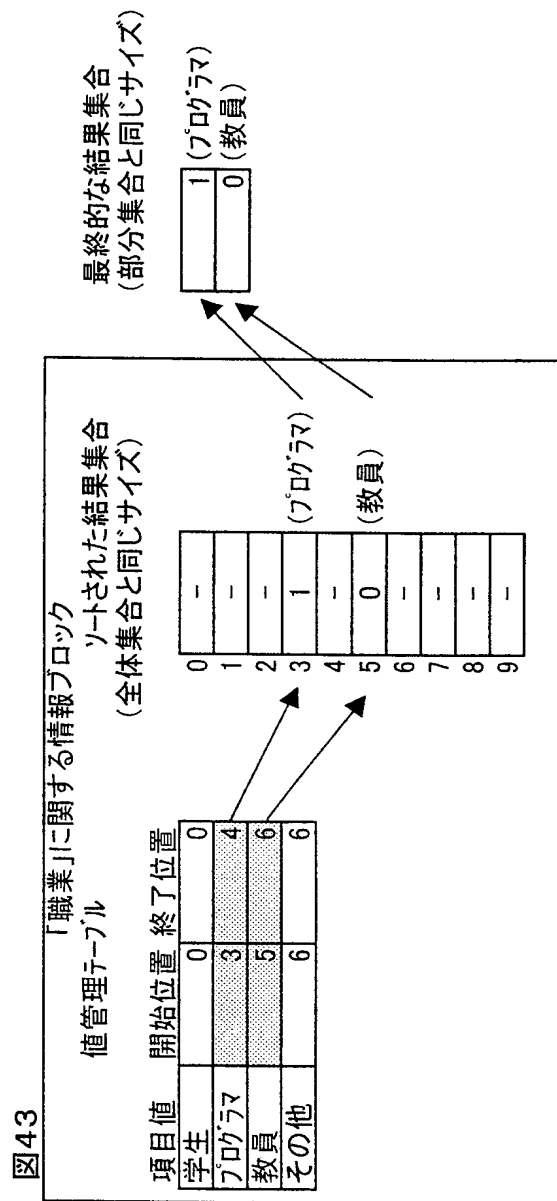
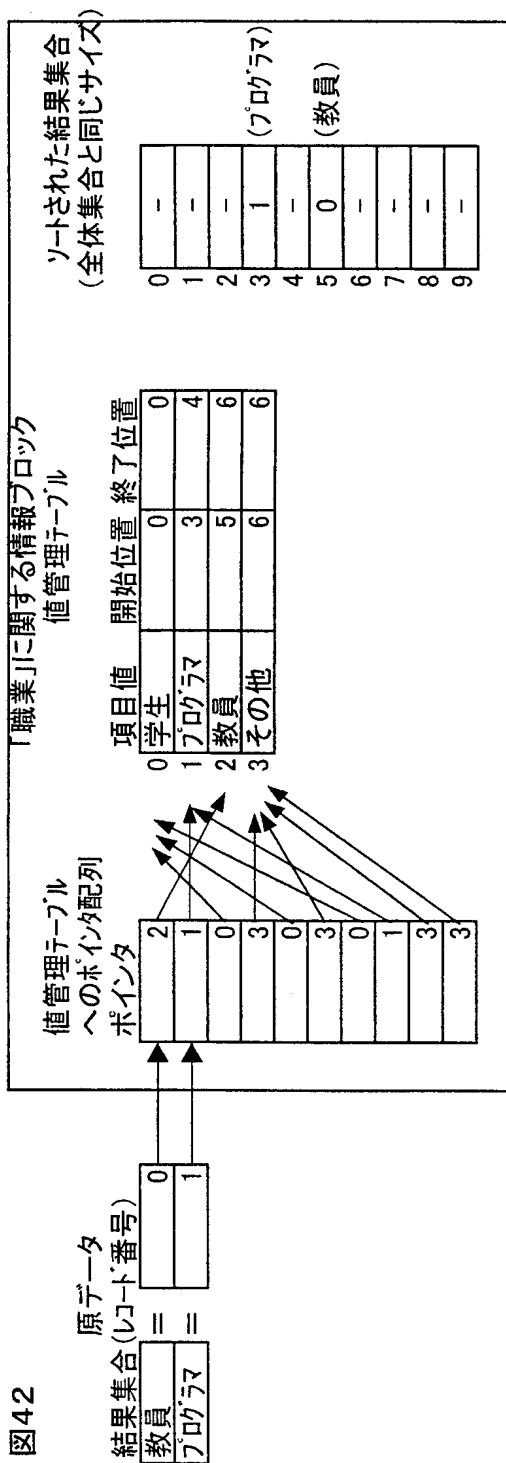


図44

レコード番号	x1万	x100	x1
000000	00	00	00
000001	00	00	01
.	.	.	.
000100	00	01	00
000101	00	01	01
.	.	.	.
010000	01	00	00
010001	01	00	01
.	.	.	.
999999	99	99	99

図45

ヒット件数	ビットフラグ	ポインタ配列	
	検索	検索	クロス集計
1000000	324	230	168
500000	260	185	83
250000	210	184	40
100000	205	127	15
10000	120	106	3
1000	35	95	0

図47

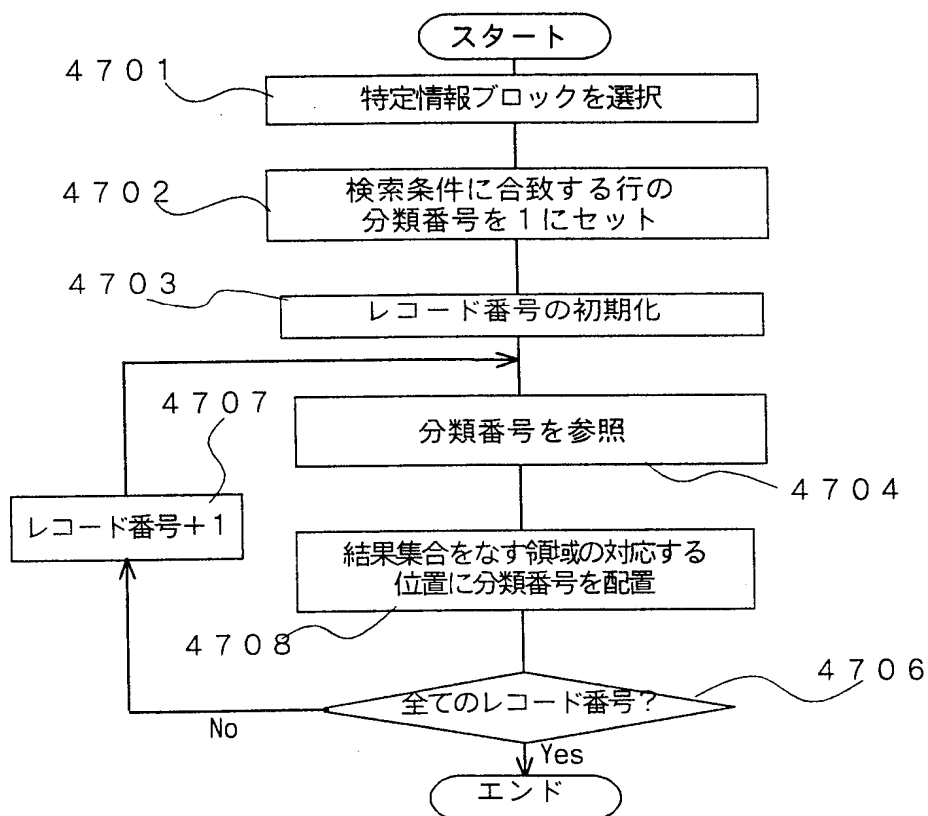


図 46

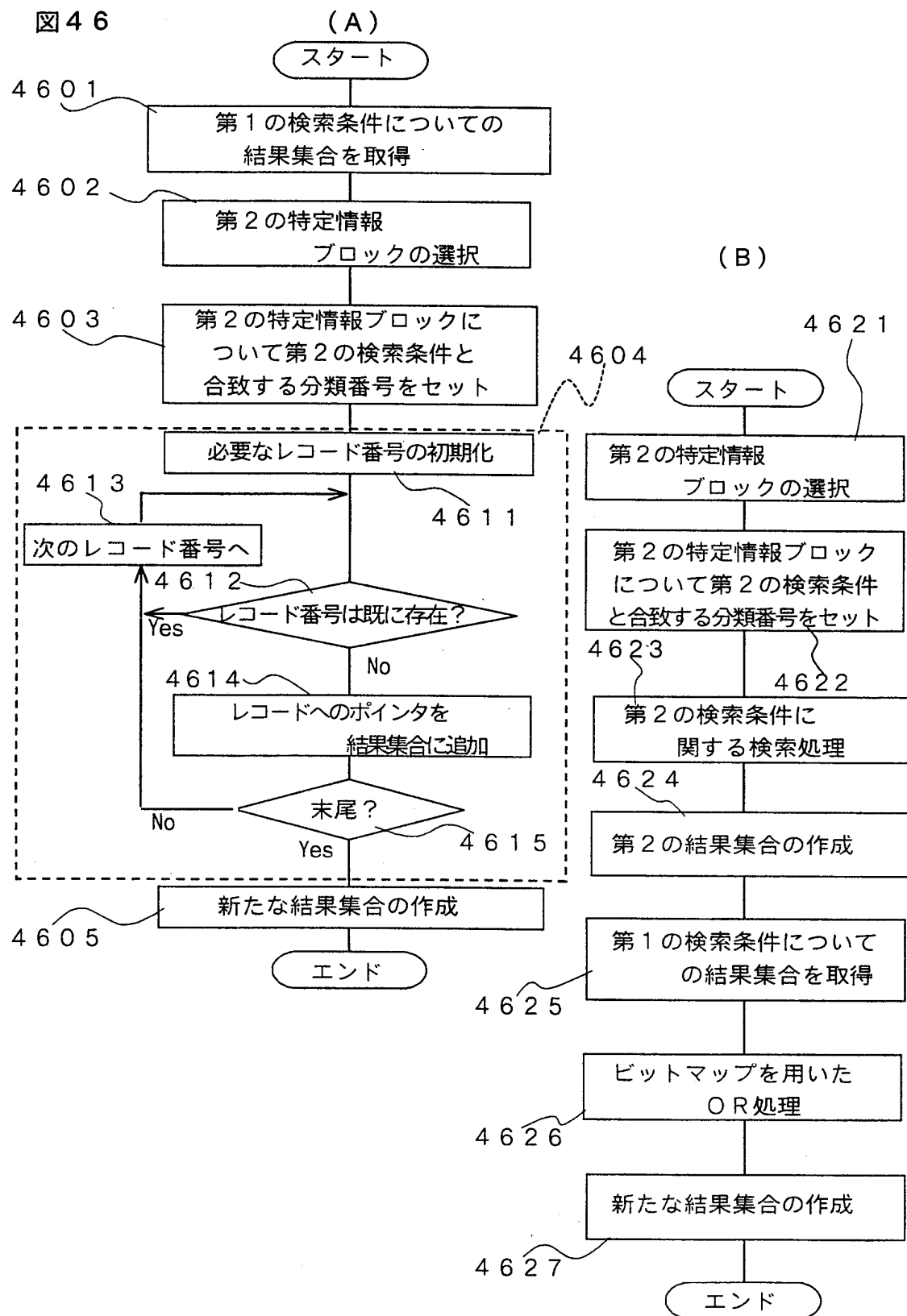


図 4 8

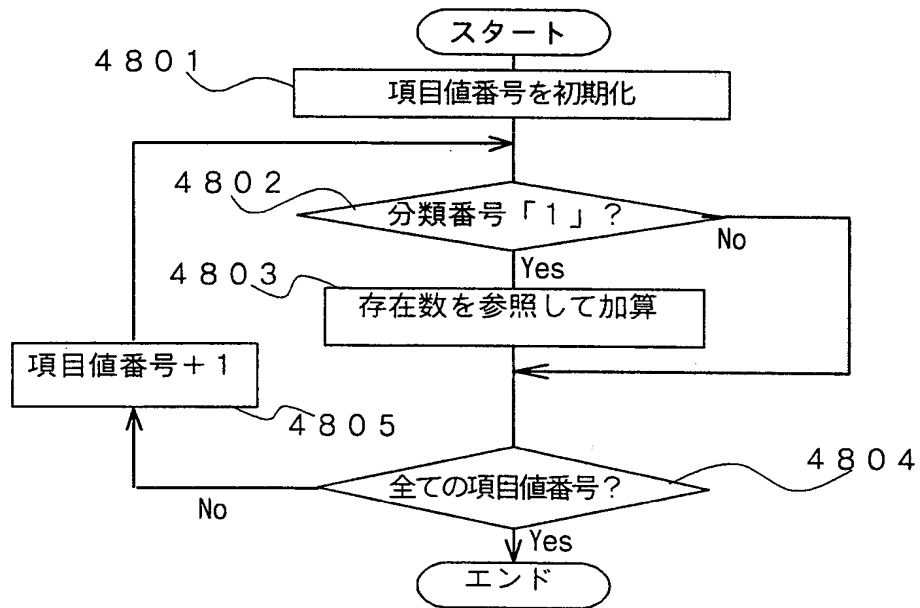


図 4 9

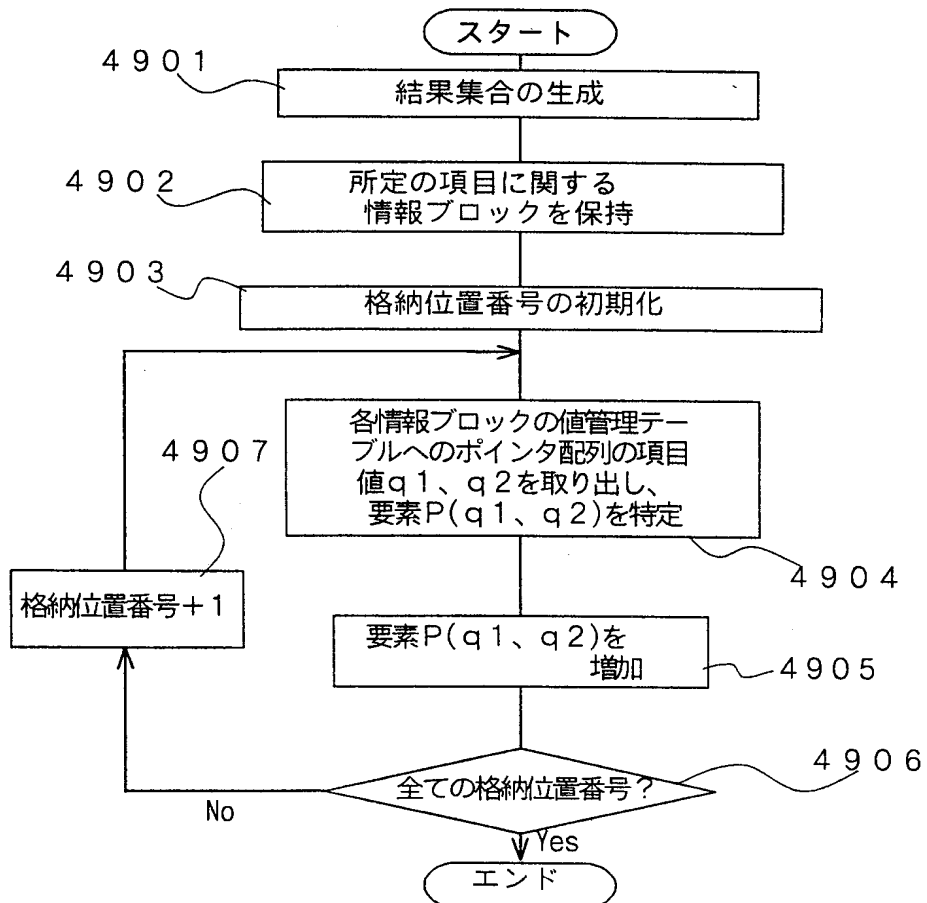
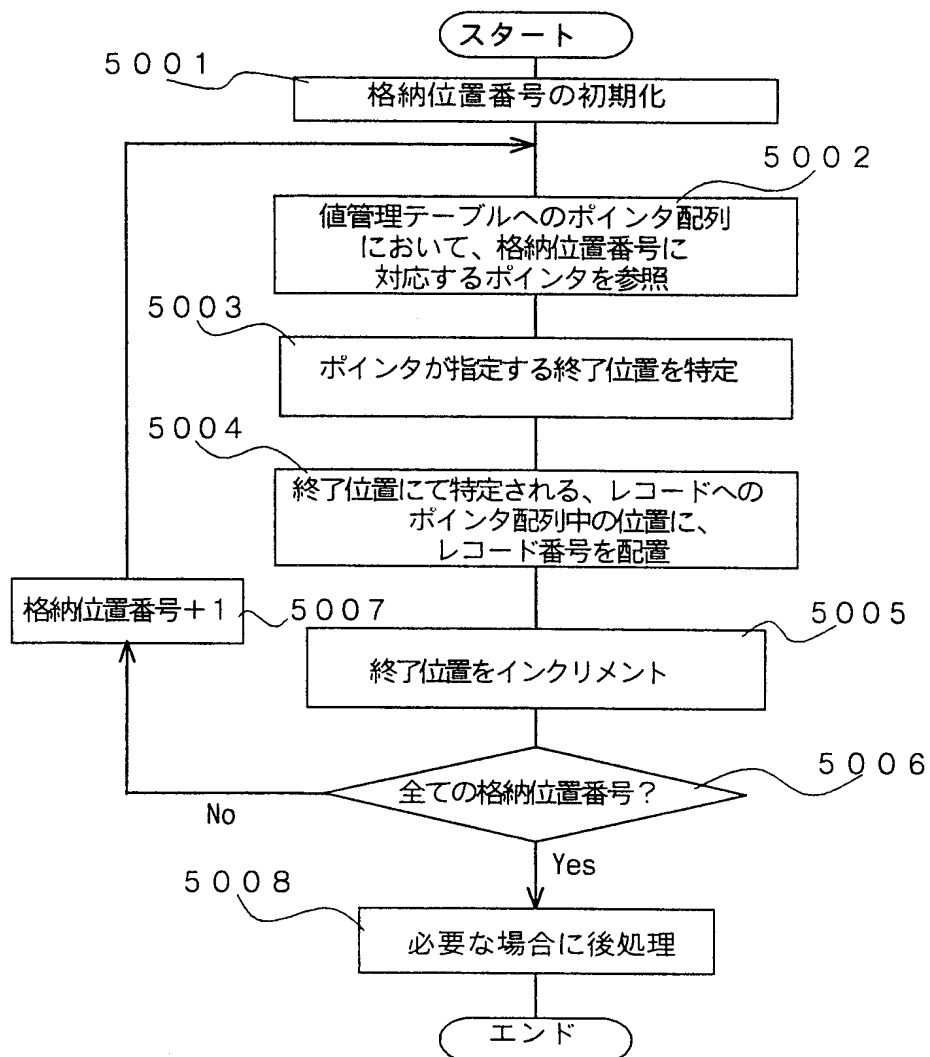


図 50



# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/JP99/04300

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> Int.Cl <sup>6</sup> G06F17/30, G06F19/00		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) Int.Cl <sup>6</sup> G06F17/30, G06F19/00		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1926-1996 Toroku Jitsuyo Shinan Koho 1994-1999 Kokai Jitsuyo Shinan Koho 1971-1999 Jitsuyo Shinan Toroku Koho 1996-1999		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) JICST File on Science and Technology (Data Ware House)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	JP, 4-128972, A (Fujitsu Ltd.), 30 April, 1992 (30. 04. 92) (Family: none)	1-3, 16, 17, 19-24
Y		8-15
X	JP, 1-219927, A (Hitachi, Ltd., et al.), 1 September, 1989 (01. 09. 89) (Family: none)	1-7, 16, 17, 19-24
Y		8-15
X	Uki, "Sybase IQ 'Dokuji data kouzou ni yoru, Data ware house he no approach'" Denshi Jouhou Tsuushin Gakkai	1-7
Y	Gijutsu Kenkyuu Houkoku (AI97-42), Vol. 97, No. 415, 2 December, 1997 (02. 12. 97), pp.51-56, particularly page 53, bitmap index	8-15
Y	Ueno, "Montage houshiki ni yoru toukei DB kensaku interface", Denshi Jouhou Tsuushin Gakkai Gijutsu Kenkyuu Houkoku (DE88-15), Vol. 88, No. 127, 22 July, 1988 (22. 07. 88), pp.25-32	8-15
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed		"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
Date of the actual completion of the international search 1 September, 1999 (01. 09. 99)		Date of mailing of the international search report 14 September, 1999 (14. 09. 99)
Name and mailing address of the ISA/ Japanese Patent Office		Authorized officer
Facsimile No.		Telephone No.



**INTERNATIONAL SEARCH REPORT**

International application No.

**PCT/JP99/04300**

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JP, 6-282578, A (Fujitsu Ltd.), 7 October, 1994 (07. 10. 94) (Family: none)	8-15

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP99/04300

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 18, 25

because they relate to subject matter not required to be searched by this Authority, namely:

The computer program products stated in claims 18, 25 are computer programs themselves, and therefore they are the ones on which no international search will be made, according to PCT Rule 42(5).

2. ☐ Claims Nos.:

because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:

because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest** ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl<sup>6</sup> G 06 F 17/30, G 06 F 19/00

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl<sup>6</sup> G 06 F 17/30, G 06 F 19/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1926-1996年
日本国公開実用新案公報	1971-1999年
日本国実用新案登録公報	1996-1999年
日本国登録実用新案公報	1994-1999年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

J I C S T 科学技術文献ファイル (データウェアハウス)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	JP, 4-128972, A (富士通株式会社) 30. 4月. 1992 (30. 04. 92) (ファミリーなし)	1-3, 16, 17, 19-24
Y		8-15
X	JP, 1-219927, A (株式会社日立製作所, 外1名) 1. 9月. 1989 (01. 09. 89) (ファミリーなし)	1-7, 16, 17, 19-24
Y		8-15

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの  
「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
「O」 口頭による開示、使用、展示等に言及する文献  
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&amp;」 同一パテントファミリー文献

国際調査を完了した日

01. 09. 99

国際調査報告の発送日

14.09.99

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)  
郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

高瀬 勤

5 L

9069

電話番号 03-3581-1101 内線 3560

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	鵜木 著「SybaseIQ『独自データ構造による、データウェアハウスへのアプローチ』」, 電子情報通信学会技術研究報告(AI97-42), Vol. 97, No. 415, 2. 12月. 1997(02. 12. 97), pp. 51-56, 特に、p. 53. ビットマップ・インデックス	1-7
Y		8-15
Y	上野 著「モンタージュ方式による統計DB検索インタフェース」, 電子情報通信学会技術研究報告(DE88-15), Vol. 88, No. 127, 22. 7月. 1988(22. 07. 88), pp. 25-32	8-15
Y	JP, 6-282578, A(富士通株式会社)7. 10月. 1994(07. 10. 94) (ファミリーなし)	8-15

## 第 I 欄 請求の範囲の一部の調査ができないときの意見 (第 1 ページの 2 の続き)

法第 8 条第 3 項 (PCT 17 条 (2) (a)) の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. ☒ 請求の範囲 18, 25 は、この国際調査機関が調査をすることを要しない対象に係るものである。つまり、  
請求の範囲 18, 25 のコンピュータプログラムプロダクトはコンピュータプログラムそのものと認められ、コンピュータプログラムは特許協力条約に基づく国際出願等に関する法律施行規則第 42 条 (5) に定められた国際調査を要しないものである。
2. ☐ 請求の範囲 \_\_\_\_\_ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. ☐ 請求の範囲 \_\_\_\_\_ は、従属請求の範囲であって PCT 規則 6.4(a) の第 2 文及び第 3 文の規定に従って記載されていない。

## 第 II 欄 発明の単一性が欠如しているときの意見 (第 1 ページの 3 の続き)

次に述べるようにこの国際出願に二以上の発明があるとこの国際調査機関は認めた。

1. ☐ 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求の範囲について作成した。
2. ☐ 追加調査手数料を要求するまでもなく、すべての調査可能な請求の範囲について調査することができたので、追加調査手数料の納付を求めなかった。
3. ☐ 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったので、この国際調査報告は、手数料の納付のあった次の請求の範囲のみについて作成した。
4. ☐ 出願人が必要な追加調査手数料を期間内に納付しなかったので、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求の範囲について作成した。

追加調査手数料の異議の申立てに関する注意

- ☐ 追加調査手数料の納付と共に出願人から異議申立てがあった。  
☐ 追加調査手数料の納付と共に出願人から異議申立てがなかった。